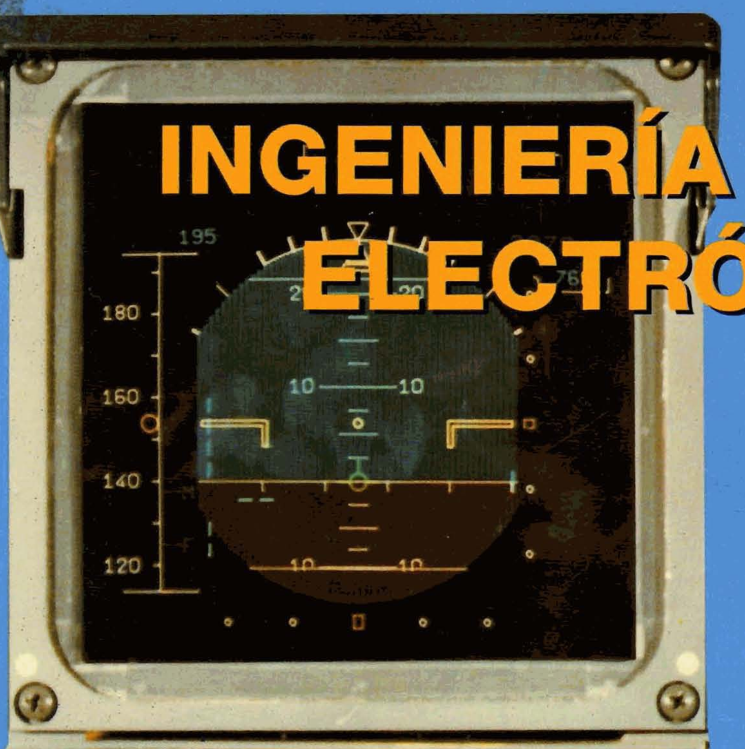
1

BIBLIOTECA TÉCNICA UNIVERSITARIA

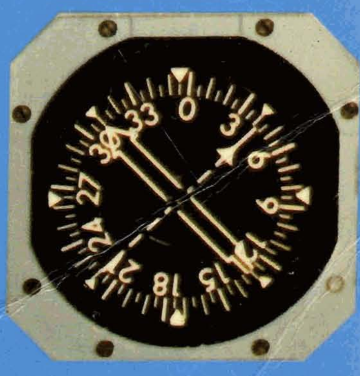
6ª EDICIÓN AMPLIADA Y ACTUALIZADA

INGENIERÍA ELECTRÓNICA







J. GONZÁLEZ BERNALDO DE QUIRÓS

 ELECTRÓNICA

J. González Bernaldo de Quirós
Catedrático de Electrónica de la Escuela Técnica Superior
de Ingenieros Aeronáuticos de la
Universidad Politécnica de Madrid

ÍNDICE

CAPITULO 1

INTRODUCCION

1.1.- Electrónica. Conceptos generales.

Electrónica es el arte de utilizar dispositivos en que se mueven cargas eléctricas para obtener, manejar y utilizar información. Los dispositivos en que se mueven las cargas eléctricas son "componentes", "circuitos y "subsistemas" o "sistemas eléctricos". La información aparece "codificada" en las variables eléctricas funciones del tiempo, como la diferencia de potencial entre dos puntos o la intensidad que circula por un conductor.

Las variables eléctricas, funciones del tiempo, que contienen información - codificada - se llaman "señales". Las componentes y circuitos manejan estas señales, y, con ellas, la información que contienen.

La Electrónica es una herramienta de cálculo. El "tratamiento de la información", "proceso de datos" y otros conceptos análogos corresponde a una idea generalizada del cálculo. La Electrónica es un eslabón del sistema de ingeniería, el que corresponde al cálculo. Por eso suele representar la parte inteligente de los sistemas .

La acertada utilización de técnicas electrónicas y no electrónicas para obtener el resultado apetecido constituye la "Ingeniería de sistemas". Normalmente las variables de entrada de un sistema y las salidas (o aplicación de resultados) no son electrónicas. En principio, puede haber sistemas que no tengan nada electrónico, si bien en la vida moderna prácticamente todos los sistemas tienen una gran parte electrónica.

1.2.- Componentes y circuitos.-

Las componentes son dispositivos eléctricos simples que relacionan las variables eléctricas relacionadas con ellas a través de ecuaciones características. Cuando estas variables contienen información, es decir, son señales, las ecuaciones de las componentes relacionan las señales.

Por ejemplo, una resistencia es una componente. La tensión eléctrica entre sus terminales e(t) y la intensidad que circula por la resistencia i(t) están relacionadas por la ley de Ohm. La ecuación de la resistencia es:

$$\frac{e(t)}{i(t)}=r$$

Los circuitos electrónicos son combinaciones de componentes que desempeñan una función determinada, generalmente una transformación de variables. Tradicionalmente, las combinaciones de componentes eléctricas se denominan **redes**, reservándose el nombre de **circuitos** para las redes **ordenadas**, que tienen una o varias entradas y una o varias salidas y cumplen una función determinada, que, en general, consiste en hacer que las salidas sean función de las entradas. La o las ecuaciones de la red definen la manipulación de la información entre las entradas y las salidas. Por ejemplo, la red de la figura 1-1 es un circuito constituido por una sola resistencia de 100 ohmios:

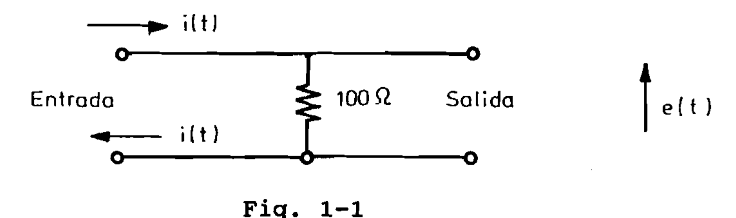


Fig. 1-1

La tensión de salida e(t) vale 100 i(t) si suponemos despreciable la intensidad de salida. Si la "información" corresponde al valor de i(t), la tensión e(t) corresponde a ese valor multiplicado por 100. El circuito es un multiplicador elemental pero rápido (instantáneo).

Los circuitos suelen ser más complicados. A veces se dividen a su vez en subcircuitos (que también se llaman circuitos) efectuando la operación lógica del circuito total por partes.

El dibujo del circuito se llama **esquema**. La figura 1-1 es el esquema de un circuito multiplicador. La correspondencia entre el circuito y el esquema es topológica; la forma del circuito real de la figura 1-1 no se parecerá al esquema y existen infinitos circuitos con el mismo esquema. No obstante, el esquema representa a todas las componentes del circuito real y sus conexiones y sirve para definir las ecuaciones del circuito (de todos los circuitos con el mismo esquema). Más adelante estudiaremos las propiedades de los circuitos y sus esquemas.

La información está contenida en las variables eléctricas (funciones del tiempo) ligadas al circuito, bien sea porque se aplican estas variables al circuito o porque se obtienen de él. Generalmente se manejan sólo dos tipos de variables, tensión e intensidad, pero es posible usar otras como cantidad de

electricidad, campo magnético, etc. La información está contenida en estas variables a través de una codificación. El código define la correspondencia entre la variable eléctrica y otra variable física del sistema que es la que se quiere manejar. Por ejemplo, la intensidad eléctrica del circuito 1-1 puede representar la temperatura en grados centígrados de un punto de un motor a razón de un amperio por grado. La tensión de salida es esa misma variable multiplicada por 100, es decir, 0,01 grados a la entrada darán 1 voltio a la salida. La salida es mucho más "sensible" que la entrada.

Las variables que contienen información se llaman **señales**. Por ejemplo, en la figura 1-1 i(t) es la señal de entrada y e(t) es la señal de salida. El modo como las señales contienen la información a través del **código**. El código de i(t) en el ejemplo citado es 1 amp./ºC y el de e(t) es 100 V./ºC. Las variables se llaman "codificadas" y la definición del código, "codificación".

Los elementos que "convierten" las señales no eléctricas en eléctricas se llaman **transductores**. Por ejemplo, un micrófono es un transductor que convierte señales acústicas (presión del aire en función del tiempo) en eléctricas. En realidad, la conversión es una codificación.

1.3.- Ingeniería electrónica. Diagramas de bloques.

El proyecto de un sistema comprende varias fases que van de lo general a lo particular.

El primer escalón técnico del proyecto (una vez superada la fase de decisión por razonamientos sociales y económicos) es la **Ingeniería de sistemas**. El ingeniero de sistemas, en contacto con el usuario, decide lo que se puede hacer para satisfacer las necesidades del consumidor y lo materializa en un esquema de principio en el que constan las partes del sistema y sus relaciones, en forma de **diagrama de bloques,** en el que cada bloque representa dentro de un rectángulo circuitos o grupos de circuitos que desempeñan una función determinada. Los bloques tienen "entradas" a las que llegan señales y "salidas" que envían las señales transformadas a otros bloques o lugares. El camino de la información (de las variables que contienen información, es decir, de las señales) está señalado por flechas que indican que este camino es irreversible, la información no vuelve al sitio de donde ha salido (o, al menos, no por el mismo camino). La información puede bifurcarse o dividirse, pero no confluir sin que esté bien definido el significado de la confluencia, por ejemplo suma o producto.

Por ejemplo, la figura 1-2 representa un esquema de dosificador automático. La tarea que el ingeniero de sistemas ha recibido del cliente y que debe realizar el sistema es echar 25 gotas de un determinado líquido en cada uno de los frascos que circulan por una cadena de producción.

El sistema funciona de la siguiente manera:

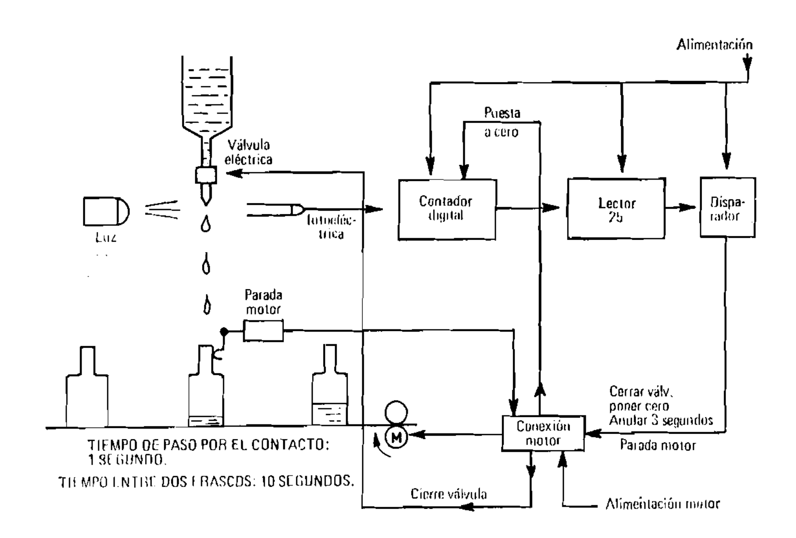


Fig. 1-2

Cada gota que cae entre la fuente luminosa y la célula fotoeléctrica interrumpe el rayo luminoso y produce un impulso eléctrico negativo (por falta de luz) a la salida de la célula. Es decir, la célula produce un nivel habitual de tensión eléctrica que se interrumpe brevemente al paso de las gotas. Los impulsos eléctricos negativos se envían por cables eléctricos (las flechas representan a estos cables o grupos de cables) a un contador electrónico, contador "digital" que "cuenta" esos impulsos dando tensiones en sus salidas que corresponden al número que cuenta. Un "lector" interpreta esas tensiones y da una señal de salida cuando la combinación de tensiones a la entrada corresponde al número 25. La tensión de salida del lector, que indica que se ha llegado a 25 provoca la maniobra de cambio de frasco.

El bloque de maniobra de cambio de frasco actúa en varios sitios. En primer lugar cierra la válvula de salida de gotas y pone el contador en cero (para comenzar la cuenta siguiente). Además pone en marcha el motor de arrastre de los frascos (cortocircuitando el contacto de parada actuado por los frascos). El cortocircuito del interruptor de parada puede estar temporizado, dando tiempo suficiente para que el frasco se libere del interruptor, por ejemplo 3 segundos. El cierre de la válvula debe durar hasta la llegada del frasco siguiente.

Cuando llega un nuevo frasco vacío a colocarse bajo el dosificador, éste actúa sobre el interruptor del motor, parándolo. También debe actuar sobre la válvula de paso de las gotas, para que vuelva a comenzar todo el proceso.

El ingeniero de sistemas va algo más lejos y determina

claramente las señales de entrada y salida de cada bloque, así como la transformación que debe realizar el bloque y su alimentación y cualquier otra condición de funcionamiento. El conjunto de estas normas se llama **especificaciones** del bloque.

Las especificaciones del bloque "lector" serán, por ejemplo, las siguientes:

<u>Entrada</u>:Cinco conexiones donde aparecen tensiones de 0 ó 12 voltios según el número del contador. Por ejemplo, el número 25 = 1x16 + 1x8 + 0x4 + 0x2 +1x1 está definido porque los conductores primero, segundo y quinto tienen tensión (12 voltios respecto a "masa" o "tierra", referencia general del sistema) y los tercero y cuarto tienen 0 voltios.

<u>Salida</u>:Al aparecer a la entrada la combinación 25, y en ninguna otra, aparecerá a la salida un "impulso", es decir, en la conexión de salida subirá la tensión de 0 a 12 voltios, permanecerá 3 segundos en 12 voltios y luego regresará a 0.

<u>Alimentación</u>: El circuito lector estará alimentado por una batería (o fuente de alimentación equivalente) de 12 voltios. El consumo máximo será de 0,2 amp.

En este sistema, el contador, el lector y todo el cableado lógico son electrónicos, compuestos por circuitos o grupos de circuitos. El resto del sistema es electromecánico.

El segundo escalón técnico lo constituye la **Ingeniería de circuitos**, que divide los subsistemas en circuitos y proyecta los circuitos a partir de las componentes disponibles.

Por ejemplo, el lector del sistema dosificador se dividirá en 3 circuitos, como se ve en la figura 1-3:

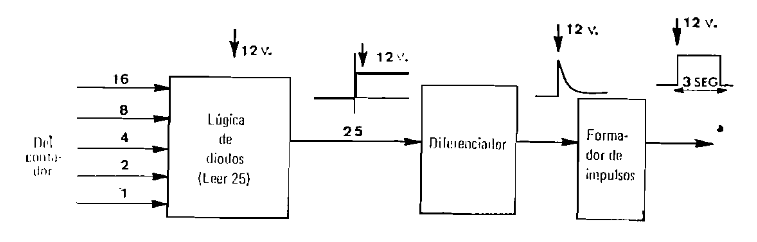


Fig. 1-3

En esta figura se definen tres circuitos que forman el lector:

<u>Lógica de diodos</u>: Es el circuito que detecta la presencia de la combinación 25 en la entrada, subiendo la salida (dando salida "alta") cuando haya un 25 a la entrada.

<u>Diferenciador</u>: Que convierte la subida de tensión en un impulso corto, que marca el inicio de la operación de cambio de frasco.

Formador de impulso-puerta: Que produce una forma rectangular de 3 segundos de duración, necesarios para que el frasco se libere del interruptor y no pare el motor hasta la llegada de otro frasco.

A continuación se proyecta cada uno de los circuitos. Por ejemplo, la lógica de diodos puede ser la de la figura 1-4, como se verá más adelante.

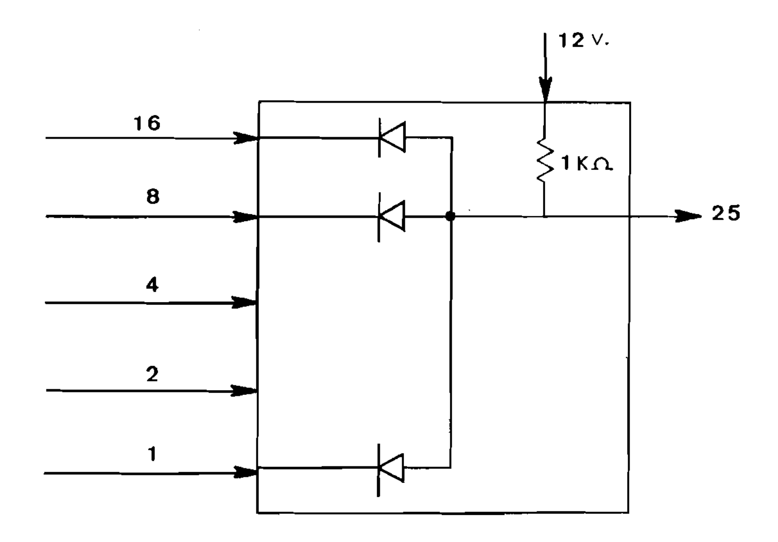


Fig. 1-4

Por último, el tercer escalón lo constituye la **Ingeniería de componentes**. En el ejemplo anterior hemos usado tres diodos y una resistencia. Cada uno de estos elementos debe ser fabricado cumpliendo determinadas especificaciones, como por ejemplo:

Valor: 1KΩ ± 10%, a 25° C.
Potencia máxima disipada: 0,25 vatios a temperatura hasta 60° C.
Condiciones ambientales: Humedad hasta 90%.Presión atmosférica.

La primera condición especifica la tolerancia, la segunda es un límite de potencia que se puede disipar en esta resistencia sin quemarla, la tercera fija algunas precauciones de fabricación, como la naturaleza del aglomerante, laca, etc.

1.4.- Aplicaciones de la Electrónica. Técnicas afines.

La Electrónica tuvo su origen en la Radiotecnia. Hoy día, las aplicaciones típicas iniciales (Comunicaciones) han pasado a ser sólo una aplicación más y la Electrónica, como parte lógica o inteligente de los sistemas se ha extendido a prácticamente todos los campos de la Ingeniería. A título de ejemplo, citaremos algunas de las aplicaciones más importantes:

Telecomunicaciones (Radiotecnia): Las ondas acústicas que

contienen la información se transforman en ondas eléctricas y éstas pasan su información ("modulan") a ondas electromagnéticas que se propagan a larga distancia. En el receptor se hace la transformación inversa, obteniendo la información.

Televisión: Por un proceso análogo al anterior, modulando ondas electromagnéticas, se transmite información visual.

Telemetría: Es el arte de enviar datos a distancia por un principio parecido a las anteriores.

Radar: Es el arte de medir distancias basándose en el tiempo de ida y vuelta de las ondas electromagnéticas a los obstáculos que las reflejan.

Radiolocalización: Es el arte de situar objetos, basándose en las medidas de campos eléctricos radiados desde focos de referencia.

Cálculo analógico: Es el arte de resolver ecuaciones diferenciales por analogía con las ecuaciones que deben satisfacer circuitos eléctricos.

Cálculo digital: Es el arte de hacer operaciones numéricas mediante circuitos electrónicos.

Automática:Es el arte de sustituir al operador humano por circuitos que realizan la misma función lógica.

Instrumentación: Es el arte de medir variables eléctricas o no por medios electrónicos.

Aviónica: Es la aplicación de las técnicas de Automática e Instrumentación a bordo del avión.

Algunas de las técnicas enumeradas pueden ser híbridas de Electrónica y otras. Por ejemplo, la Automática puede usar elementos electrónicos o hidráulicos o neumáticos, solos o combinados.

1.5.- Estudio de la Electrónica.- Plan general del libro.

Comenzaremos por el estudio de **señales** y de los métodos matemáticos para su tratamiento. A continuación pasaremos a estudiar **componentes, redes y circuitos**, tanto los de función analógica como los digitales.

Por último estudiaremos **sistemas electrónicos** de carácter diverso, tratando de cubrir el espectro de las aplicaciones más importantes.

CAPITULO 2

METODOS DE ANALISIS DE LAS SEÑALES

2.1.- Introducción.

En el capítulo anterior hemos visto que los circuitos electrónicos son operadores que actúan sobre las variables eléctricas, funciones del tiempo, transformándolas. La información está contenida en la forma de variación con el tiempo de estas variables, generalmente tensión e intensidad, de la misma forma que la información que se transmite en un lenguaje hablado está contenida en la variación con el tiempo de las variables acústicas, presión y velocidad del aire en un punto. En el ejemplo del lenguaje hablado, para extraer la información (para entenderlo) es necesario conocer el código o sistema con el que está contenida la información en las variables acústicas. Este código constituye un idioma. En nuestro caso el código está inventado en cada caso de acuerdo con el tipo de información.

Se llama **señal** a la variable que contiene información, distinguiéndola así de otras variables auxiliares que son necesarias para el funcionamiento del circuito, pero que no contienen información (por ejemplo, la tensión de alimentación). El sistema por el cual se introduce la información en la señal es la **codificación**. El conjunto de reglas para realizar la codificación constituye el **código**.

2.2.- Tipos de señal de uso normal en Electrónica.

Entre las infinitas formas de representar la información, se suele codificar de la forma más sencilla y directa para la información de que se trate. Las señales de uso normal en Electrónica pueden clasificarse en tres grandes grupos, que corresponden también a tres tipos distintos de información :

a) Señales periódicas. Se entiende por tales aquellas cuyo valor se repite cada vez que transcurre un período. No basta con que estén definidas en un intervalo o período, tal como se definen

en matemáticas, porque en un sistema real aparecen siempre efectos de valores exteriores al período, que deben estar definidos por la **repetición.**

En la realidad no existe ninguna función periódica que corresponda a esta definición, porque, evidentemente, no existe ninguna función que se repita desde tiempo menos infinito a mas infinito.Para nosotros es suficiente que los regímenes transitorios de inicio y final de la señal no tengan significado ni influyan en el funcionamiento de los circuitos durante la zona de tiempo en que nos movemos.

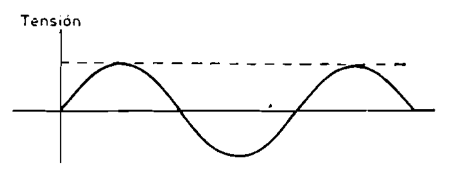
Toda la información de una onda periódica está contenida en un solo período, o en cada período, ya que los demás períodos son una repetición

Estas señales corresponden por naturaleza a las **ondas** de todo tipo, sonoras, electromagnéticas, etc.

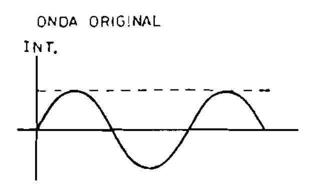
El caso particular más simple es el de las ondas sinusoidales al que, además, se pueden referir las no sinusoidales por descomposición en serie de Fourier.

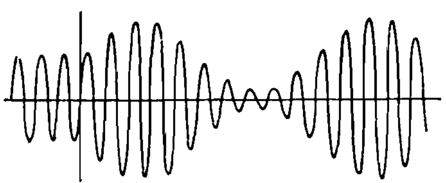
El modo más simple de representar una señal sinusoidal es con otra señal sinusoidal de la misma frecuencia y amplitud proporcional a la señal original,con un factor de escala.

PRIMERA REPRESENTACION



SEGUNDA REPRESENTACION (AM)





TERCERA REPRESENTACION (FM)

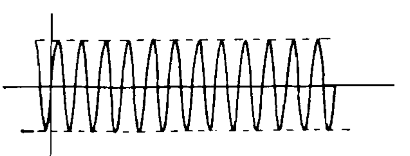


Fig. 2-1.

Si, por ejemplo, se trata de representar un tono musical, caracterizado por una frecuencia (o tono) y una intensidad (en relación con una unidad de referencia), podemos representarla, como se ve en la figura 2-1, de una de las tres formas siguientes:

Por una onda eléctrica de la misma forma y frecuencia y cuya amplitud sea equivalente a la intensidad sonora a través de un factor de escala.

Por una onda eléctrica sinusoidal de alta frecuencia cuya amplitud varíe a la frecuencia de la onda original y con un tanto por ciento de variación proporcional a la amplitud de la onda original.(**Modulación en amplitud.**)

Por una onda eléctrica sinusoidal de alta frecuencia cuya desviación de frecuencia ocurra a la frecuencia de la onda original, con desviación de frecuencia proporcional a la amplitud de la onda original.(**Modulación en frecuencia.**)

En los dos últimos casos la frecuencia de la onda modulada no tiene significado alguno relacionado con la onda original y puede ser cualquiera, o tener otro significado como puede ser la identificación de la procedencia del tono musical.

El ejemplo anterior sirve para explicar el hecho de que puede haber elementos mezclados con la señal y que no guarden relación con la información codificada, aunque sí con otras fuentes de información. La figura 2-2, que mantiene una codificación directa y sencilla de una onda sinusoidal, explica por sí misma la posibilidad de añadir un nivel constante o incluso de mezclar otras señales, en forma de ondulaciones de mayor frecuencia sin que esto perturbe para nada la inteligencia de la información, siempre que en el código esté suficientemente claro que estos elementos, perfectamente identificables, no contienen información.

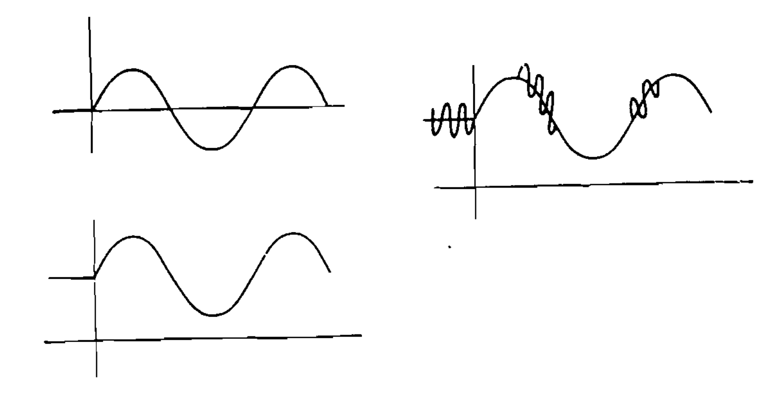


Fig. 2-2

b) Señales que representan variables funciones cualesquiera del tiempo.

Por ejemplo, una temperatura o una presión funciones del tiempo. Estas señales son las más generales, incluyendo todas las otras. Estas señales reciben el nombre de **analógicas**, por ser el tipo usado en los calculadores analógicos que se basan en el parecido de las ecuaciones de dos fenómenos físicos distintos para sustituir el cálculo de uno por el del otro.

La codificación de estas señales es sencilla, representándose normalmente por señales eléctricas proporcionales.

Sin embargo, debido a la generalidad de esta definición, el tratamiento matemático de estas señales es difícil por lo que normalmente se recurre a los numerosos recursos de cálculo operacional para llevarlas a otros tipos de señales, sinusoidales por ejemplo, de tratamiento eléctrico más conocido.

c) Funciones que sólo toman valores discretos con el tiempo, o que sólo tienen significado en momentos determinados.

Este tipo de señales son las típicas de la información numérica en todas sus formas. El cambio entre valores fijos de una variable se llama **conmutación**, que puede ocurrir también en momentos determinados, regularmente espaciados. Por ejemplo, podemos imaginar un sistema de codificación binario en que la tensión cero voltios (o cualquier tensión inferior a cero voltios) represente el número **cero** y la tensión 12 voltios (o cualquier tensión superior a 6 voltios) represente el número **uno**. Además, podemos convenir en dividir el tiempo en intervalos fijos, por ejemplo, de un milisegundo, y dar significado a los valores de la tensión en esos momentos. La señal de la figura 2-3 representaría entonces la secuencia 011001, que en base 2 sería el número 25. Se dispone de un gran margen de deformación de señal sin cambiar de significado y la figura inferior también representa el mismo número.

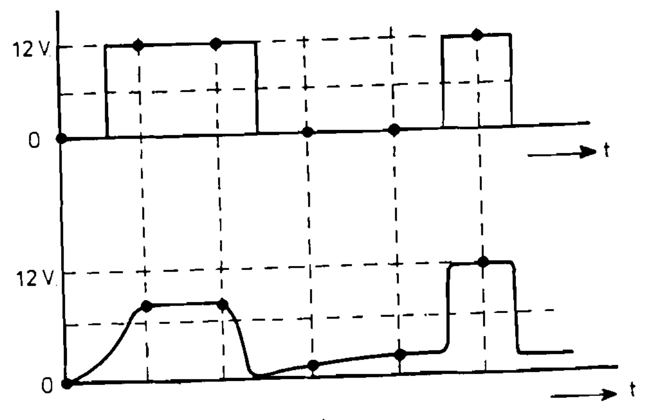


Fig. 2-3

Cuando se toman valores continuos (analógicos) en momentos determinados (por ejemplo, valor de una variable cada segundo), se trata de un tipo especial de conmutación que se llama **muestreo**.

2.3.- Análisis de las señales.

Las señales son variables eléctricas, medidas en un punto del sistema, que contienen información. Las señales, como portadoras de información, se envían de un punto a otro del sistema, pueden bifurcarse o componerse entre sí, de forma que la información de la señal resultante sea la combinación deseada de las informaciones de las señales componentes,se amplifican de forma que sean fácilmente legibles, se presentan en dispositivos de lectura, etc...Todas estas operaciones de transformación de señales son el objeto de la Electrónica.

Sin embargo, antes de hacer un estudio detallado de los circuitos que llevan a cabo estas labores, es necesario hacer un análisis de las señales mismas, análisis al que dedicaremos el resto de este capítulo.

Es necesario aclarar el concepto de que el análisis de las señales no supone ningún cambio real de las señales, ya que no hay ningún cambio en una señal cuando, por ejemplo "se la descompone" en suma de otras dos. La descomposición es una entelequia matemática que no se nota en la realidad, ya que la señal aparecerá en su totalidad y no como suma de componentes. Sin embargo, a través de posteriores operaciones o manipulaciones de la señal, pueden ponerse de manifiesto las componentes o sus propiedades.

La naturaleza de nuestro análisis debe ir ligada al tipo de operaciones que luego seamos capaces de realizar. Por ejemplo, la descomposición de una función del tiempo en funciones ortogonales sinusoidales es interesante porque luego veremos que somos capaces de manejar especialmente bien las funciones sinusoidales, eliminando algunas componentes (filtrando), amplificando unas más que otras (amplificación selectiva), etc...

A continuación analizaremos las señales descritas en el párrafo anterior, bajo los epígrafes a) y b), dejando para otro lugar el estudio de las variables de valores discretos, "digitales", cuyo tratamiento es distinto de el de las señales periódicas o analógicas.

2.4.- Análisis de las señales periódicas. Desarrollo de Fourier.

Todas las funciones del tiempo cuyo valor interesa en un intervalo determinado, pueden sustituirse en ese intervalo por su desarrollo en serie de Fourier, si se cumplen ciertas condiciones matemáticas que se encuentran casi siempre en las funciones que representan fenómenos físicos. La suma de los términos del desarrollo representa una función que coincide con la original en el intervalo y que es periódica fuera de éste, teniendo por período el intervalo elegido. Si la función original

es periódica, el desarrollo de Fourier es también válido fuera del intervalo. La utilización del desarrollo en serie de Fourier como instrumento de descomposición de funciones periódicas en suma de componentes sinusoidales tiene una utilización muy extensa en Electrónica, debido al tradicional conocimiento en profundidad de las funciones sinusoidales.

El desarrollo de Fourier es sólo un caso particular de descomposición de una función en componentes "ortogonales", cuyo estudio completo se sale del propósito de este libro, debiendo consultarse la literatura especializada en su caso. Es preciso señalar que aunque el desarrollo es válido para funciones definidas en un intervalo (el período) y cuyo valor en el resto del tiempo no tiene interés matemático, en la práctica sólo lo aplicaremos a funciones periódicas, cuyo valor se repite cuando transcurre un período y que coinciden con la suma de Fourier en todo el tiempo, y no solamente en el intervalo, debido a que en los sistemas reales no se puede ignorar que la misma función puede aparecer en otros puntos del sistema retrasada o adelantada y, en ese caso, hay que tener en cuenta también los valores fuera del intervalo. Por ello sólo estudiaremos el desarrollo de Fourier de funciones **periódicas** y no simplemente definidas en un intervalo.

Inmediatamente surge la duda de que en la práctica no existen las funciones periódicas, ya que parece que todos los fenómenos reales tienen un principio y un fin. No obstante, es suficiente aproximación el suponer que una función es periódica cuando el principio y el fin, es decir, los regímenes transitorios están suficientemente alejados, y , para efectos de aplicación práctica, la función se extiende de menos infinito a más infinito.

En primer lugar, el desarrollo de Fourier de la función periódica f(t) puede escribirse en la forma:

$$f(t) = a_0 + a_1 \cos \omega t + a_2 \cos 2\omega t + a_3 \cos 3\omega t + \dots$$
$$+ b_1 \operatorname{sen} \omega t + b_2 \operatorname{sen} 3\omega t + b_3 \operatorname{sen} 3\omega t + \dots$$

siendo ω la **pulsación** de la onda periódica, es decir $\omega = 2 \pi f$ radianes por segundo, siendo f la **frecuencia** de la onda en ciclos por segundo o "Hertzios", que, a su vez, es el inverso del período o intervalo del desarrollo **T** en segundos.

Insistimos una vez más en que la sustitución de f(t) por su desarrollo es válida dentro del intervalo T si la función no es periódica y válida siempre si f(t) es periódica de período T, lo que supondremos siempre en adelante.

Los términos del desarrollo pueden agruparse por frecuencias, haciendo la transformación :

$$a_n \cos n\omega t + b_n \operatorname{sen} n\omega t = A_n \cos (n\omega t + \phi_n)$$

con lo que el desarrollo queda :

$$f(t) = a_0 + A_1 \cos(\omega t + \phi_1) + A_2 \cos(2\omega t + \phi_2) + \ldots$$

Del mismo modo podríamos haber transformado el desarrollo en uno de senos (con su fase). Convenimos en utilizar el desarrollo en cosenos (con sus fases).

La función **del tiempo** $f(t)$ puede sustituirse por la suma de :

a) Un valor medio $a_0 = A_0$ que llamaremos **componente continua** de $f(t)$.

b) Una onda periódica de frecuencia igual a la de $f(t)$, amplitud A_1 y fase ϕ_1, que llamaremos **fundamental.**

c) Una serie infinita de ondas de frecuencia doble, triple, etc... de la de $f(t)$, amplitudes A_2, A_3, A_4 ... y fases ϕ_2, ϕ_3, ϕ_4 ... que llamaremos **armónicos de orden 2, 3, 4 ...** y cuya amplitud debe tender a cero con el orden ya que el desarrollo de Fourier, si existe, es convergente.

Algunas propiedades que conviene recordar son:

1) Si la función $f(t)$ es simétrica respecto al origen, es decir, si $f(t) = f(-t)$, el desarrollo inicial sólo contiene cosenos y en el desarrollo de cosenos todos los desfases son cero.

2) Si además existe una simetría dentro del período tal que $f(t) = - f(t + T/2)$, el desarrollo sólo contiene cosenos de múltiplos impares de la frecuencia de $f(t)$.

3) El desarrollo, en las dos formas descritas, es único.

4) Cada coeficiente del desarrollo representa físicamente el "contenido" en $f(t)$ de la componente sinusoidal correspondiente. Es decir, la componente $a_n \cos(n\omega t + \phi_n)$ es la función de esa forma que **más se parece** a $f(t)$. Matemáticamente esto quiere decir que los coeficientes a y ϕ de este término hacen mínima la desviación media cuadrática entre el término y $f(t)$.

5) Los coeficientes del desarrollo se calculan por las fórmulas:

$$a_0 = \frac{1}{T}\int_{t_0}^{t_0+T} f(t)\,dt$$

$$a_n = \frac{2}{T}\int_{t_0}^{t_0+T} f(t)\cos n\omega t\,dt$$

$$b_n = \frac{2}{T}\int_{t_0}^{t_0+T} f(t)\, sen n\omega t dt$$

donde el "intervalo" del desarrollo es el tiempo T entre t_0 y t_0+T, ω es la pulsación $2\pi f$.

Estas fórmulas se demuestran fácilmente si se multiplican ambos miembros de la primera fórmula del desarrollo por cos $n\omega t$ o sen $n\omega t$ y se integra a lo largo de un período. Todos los términos se hacen cero excepto el enésimo, ya que:

$$\int_{t_0}^{t_0+T} sen j\omega t \cos k\omega t dt = 0$$

y

$$\int_{t_0}^{t_0+T} sen j\omega t\, sen k\omega t dt$$

o

$$\int_{t_0}^{t_0+T} \cos j\omega t \cos k\omega t dt$$

valen cero si j es distinto de k y T/2 si j=k.

Los coeficientes del desarrollo en cosenos se hallan de la igualdad:

$a_n \cos n\omega t + b_n \text{ sen } n\omega t = A_n \cos (n\omega t + \phi_n)$

Desarrollando el segundo miembro e identificando los coeficientes:

$a_n = A_n \cos \phi_n$
$b_n = - A_n \text{ sen } \phi_n$

de donde:

$tg\ \omega_n = - b_n / a_n$
$A_n = (a_{n2} + b_{n2})^{1/2}$.

2.5.- Espectro de una señal periódica. Tiempo y frecuencia.

El desarrollo de Fourier en cosenos equivale a un cambio de variable independiente de la función f(t). A cada valor de n, es decir , a cada valor de la pulsación $n\omega$ o de la frecuencia nf (ω misma, la pulsación, no es una variable, sino un parámetro que actúa como factor constante), incluidos los valores "0" y 1, correspondientes al valor medio y al término fundamental, corresponde un valor de la amplitud A y otro de la fase ϕ (cero en el caso n=0). Si no se conoce f(t), pero se sabe que es periódica de pulsación ω y se conocen las funciones discretas A y ϕde n, se puede reproducir f(t) escribiendo el desarrollo y

realizando la suma de la serie resultante. Tomando como variable la pulsación $n\omega$, podemos definir una **correspondencia operacional** entre la función del tiempo f(t) y las funciones A y ϕde $n\omega$. Para una representación más clara de la variable pulsación podemos cambiar la denominación de la pulsación de f(t) a ω_0, parámetro, y reservar la denominación de ω para la pulsación genérica, hasta ahora $n\omega$.

El conjunto de las funciones A y ϕde ω que corresponde a una función f(t) se llama **espectro** de f(t). El espectro está constituido por dos funciones de la pulsación (o de la frecuencia) y **no** son funciones del tiempo, que ha desaparecido. En la correspondencia biunívoca entre f(t) y su espectro se ha cambiado la variable tiempo por la variable frecuencia. En el lenguaje operacional diremos que hemos pasado del "dominio" del tiempo al de la frecuencia.

Se comprende que , en las funciones periódicas, el tiempo, variable por excelencia para la descripción de los fenómenos secuenciales (con principio, transcurso, fin, orden) pierde su carácter esencial. La representación misma (incluso en el plano del tiempo) de los fenómenos periódicos confunde los preceptos de pasado, presente, futuro. En una función periódica es fácil predecir el futuro, mera repetición del presente y del pasado. En lugar del tiempo, adquieren importancia las frecuencias, las repeticiones por período, reflejadas en los armónicos.

Los valores de A en función de ω se llaman **rayas espectrales**, según denominación originada en Física. Las funciones periódicas presentan rayas espectrales distanciadas ω_0, pulsación de f(t) o f (frecuencia) si se representan en función de la frecuencia nf. Si el período es largo, la pulsación es pequeña y las rayas se aproximan entre sí. Si el período tiende a infinito (cualquier función no periódica puede considerarse periódica de período infinito), las rayas se juntan y el espectro se hace continuo.

Más adelante volveremos sobre el espectro de las funciones no periódicas, pero, por ahora, puede insistirse en que el espectro de las funciones periódicas es discontinuo.

2.6.- Espectro de la función a cos ω_0t.

El desarrollo en serie de Fourier de esta función es ella misma. El espectro se compone de un solo punto de valor a para la pulsación ω_0 siendo nulo todo el resto. También es nulo el espectro en fase. En la figura 2-4 está representada la función y su espectro.

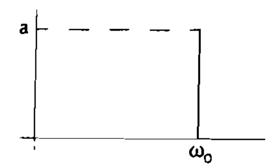


Fig. 2-4

2.7.- Espectro de la función a sen $\omega_0 t$.

La función es $f(t) = a \operatorname{sen} \omega_0 t = a \cos(\omega_0 t - \pi/2)$. El espectro se compone de un solo punto en cada diagrama. El espectro en amplitud es el mismo del caso anterior. El espectro en fase tiene un punto colocado a $-\pi/2$ en la pulsación ω_0. De una forma más general, el espectro de:

$$f(t) = a \operatorname{sen}(\omega_0 t + \phi_0) = a \cos(\omega_0 t + \phi_0 - \pi/2)$$

se compone también de un punto en cada diagrama , como se ve en la figura 2-5.

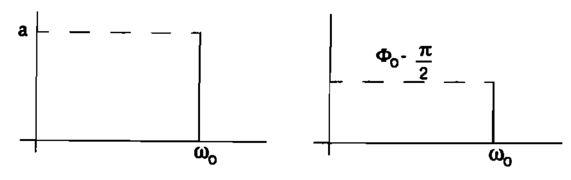


Fig. 2-5

El espectro en amplitud es el mismo del caso anterior. Un desplazamiento de f(t) en el tiempo (un retardo o adelanto de fase) no cambia el espectro en amplitud, pero sí el de fase, que cambia en los valores de sus rayas espectrales.

2.8.- Espectro de la función rectangular periódica.

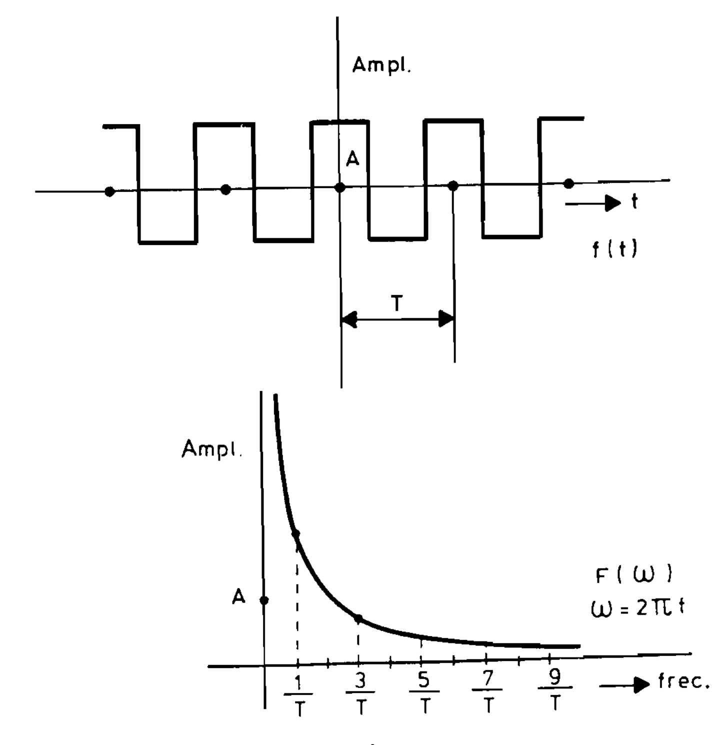


Fig. 2-6

Tomemos la función rectangular de la figura 2-6, de amplitud A y período T.

Escribiremos el desarrollo en serie de Fourier sin demostrarlo. El lector interesado en la demostración deberá referirse a la literatura matemática correspondiente. También es preciso señalar que existen numerosos formularios y manuales que contienen tablas de desarrollos de las funciones más típicas como ésta.

El desarrollo de f(t) es :

$$f(t)=\frac{4A}{\pi}\cos\omega t-\frac{4A}{3\pi}\cos 3\omega t+\frac{4A}{5\pi}\cos 5\omega t-\ldots$$

El espectro en amplitud es el de la figura. Su forma es hiperbólica, como se desprende del desarrollo. Las fases son todas cero si se escribe el desarrollo como lo hemos hecho, lo que no es convencional aunque sí habitual. Si se escribiese el desarrollo con todos los signos positivos, lo que sería más riguroso, las fases serían alternativamente cero y π, ya que cambiar el signo del coseno equivale a cambiar π la fase.

2.9.- Espectro de la función triangular.

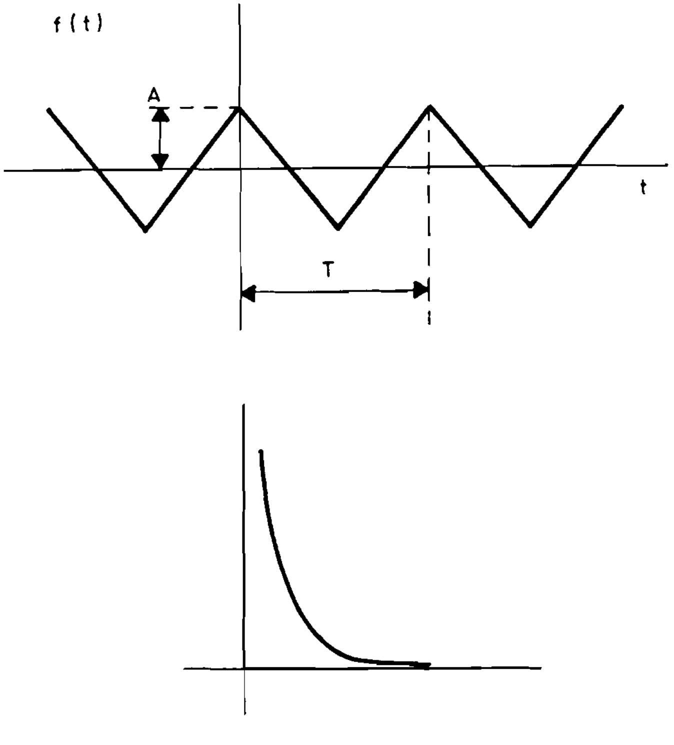


Fig. 2-7

La función triangular de amplitud A y período T, representada en la figura 2-7 tiene el siguiente desarrollo :

$$f(t)=\frac{8A}{\pi^2}\cos\omega t+\frac{8A}{9\pi^2}\cos 3\omega t+\frac{8A}{25\pi^2}\cos 5\omega t+\ldots$$

El espectro es también de tipo hiperbólico, pero de segundo grado.

El espectro está también representado en la misma figura 2-7 y es más convergente que el de la función rectangular periódica. Podríamos decir que la onda triangular "se parece más" a la función sinusoidal (onda fundamental del desarrollo) que la rectangular, ya que los armónicos de orden superior al primero son inferiores respecto al fundamental.

2.10.- Espectro de la función de impulsos rectangulares periódicos.

Tomemos la función de la figura 2-8, compuesta de "impulsos rectangulares" de ancho T y altura 1 que se repiten con el período T_r. El desarrollo es :

$$f(t)=\frac{T}{T_r}+\frac{2T}{T_r}\left[\frac{sen\pi\frac{T}{T_r}}{\pi\frac{T}{T_r}}\cos\omega_0 t+\frac{sen\frac{2\pi T}{T_r}}{\frac{2\pi T}{T_r}}\cos 2\omega_0 t+\ldots\right]$$

siendo $\omega_0 = 2\pi/T_r$ la pulsación. El espectro es la representación del término n-simo del desarrollo en función de la pulsación de ese término, $\omega = n\omega_0$. Es decir, que en abscisas se tomará ω y en ordenadas

$$\frac{2T}{T_r}\frac{sen\,n\pi\frac{T}{T_r}}{n\pi\frac{T}{T_r}}$$

y será de la forma sen x/x ($x = n\pi T/T_r$). Esta función se llama **función de muestreo** y se anula en los puntos en que $x = K\pi$. El valor de la frecuencia en esos puntos es

$nf_0 = n/T_r = x/\pi T = K/T$, es decir, 1/T, 2/T, 3/T ...

El espectro está representado en la figura 2-8.

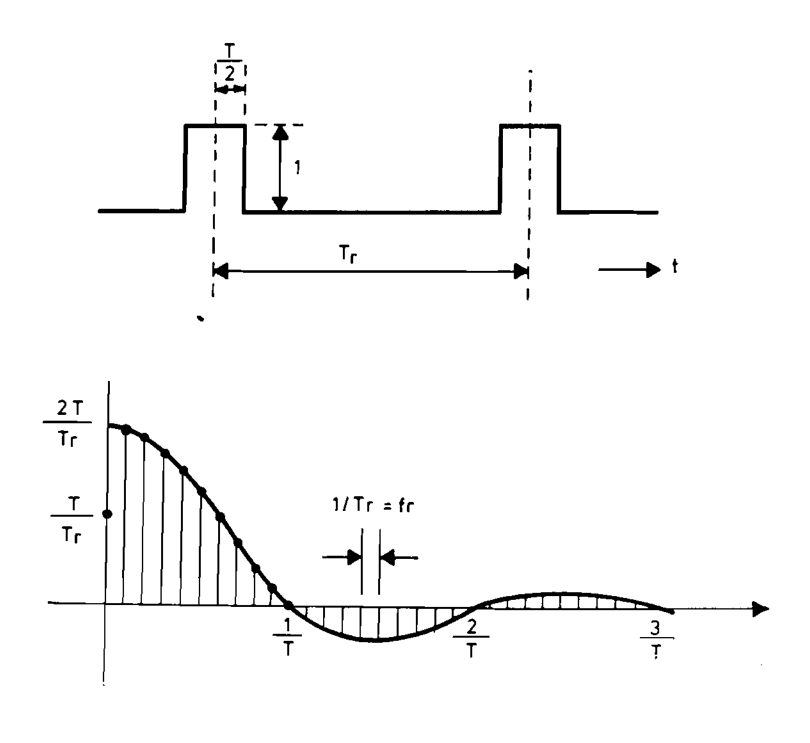


Fig. 2-8

2.11.- Espectro de la suma de funciones.

La transformación que da lugar al espectro es lineal y el espectro de la suma de funciones es la suma de los espectros. Si las funciones son periódicas y de distinto período aparecerán puntos (rayas) en las frecuencias o pulsaciones múltiplos de cada una de las frecuencias de los sumandos. Normalmente se verán entonces varias envolventes o curvas de espectro que corresponden a los puntos de las diferentes señales, sin que tengan relación entre sí.

Al hacer esta operación se ha generalizado ya el concepto de espectro como resultado de una manipulación operacional de las señales periódicas. La señal suma de varias no es necesariamente una señal periódica, sólo lo es si las diferentes señales tienen frecuencias con una relación racional, en cuyo caso la señal resultante es también periódica, de frecuencia el máximo común divisor de las frecuencias componentes. En ese caso se dice que las señales son **coherentes** , lo que físicamente significa que proceden de una misma fuente. Pero puede ocurrir que las señales no lo sean y al sumarlas se obtenga un espectro que ya no obedece a la sencilla interpretación del desarrollo de Fourier, pero que sigue siendo válido para todas las aplicaciones que vamos a usar.

2.12.- Espectro de la función cos ω_1t cos ω_2t.

Descomponiendo el producto de cosenos en suma se obtiene:

$$\cos\omega_1 t\cos\omega_2 t=\frac{1}{2}\cos(\omega_1+\omega_2)t+\frac{1}{2}\cos(\omega_1-\omega_2)t$$

Se obtienen pues las pulsaciones (o frecuencias) suma y diferencia con la mitad de la amplitud producto de las de los factores.

Si interpretamos el producto como ω_2 " modificando" a ω_1 y suponemos $\omega_1 >> \omega_2$, observamos que el efecto de multiplicar la onda de pulsación ω_1 por la onda de pulsación ω_2 es dividir aquella en dos nuevas ondas de amplitud semiproducto y desplazadas la magnitud ω_2 por arriba y abajo de ω_1 , como se ve en la figura 2-9. Esta operación corresponde a la que llamaremos más adelante **modular** la onda ω_1, que se llamará **portadora** con la onda ω_2, que se llamará **moduladora.** La aparición de las dos nuevas componentes alrededor de la de alta frecuencia se llamará **desdoblarse.** Diremos pues que la onda ω_2 se ha desdoblado sobre la ω_1.

Hay que hacer notar que en esta operación de multiplicación **han desaparecido** las frecuencias o pulsaciones iniciales, quedando ahora solamente las suma y diferencia a las que llamaremos después **bandas laterales.**

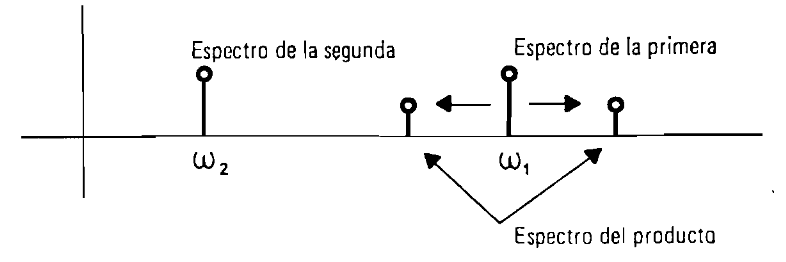


Fig. 2-9

2.13.- Espectro de la función f(t) cos ωt.

La función f(t) cos ωt tiene especial interés por intervenir en los procesos de modulación con una función cualquiera y otros diversos. Hay que distinguir el caso más frecuente de que $\omega>>\omega_0$ y el caso de que $\omega<< \omega_0$ que tiene algunas aplicaciones, siendo ω_0 la pulsación de f(t).

Escribamos el desarrollo en serie de Fourier de f(t):

$$f(t) = A_0 + A_1 \cos(\omega_0 t + \phi_1) + A_2 \cos(\omega_0 t + \phi_2) + \ldots$$

Supongamos que $\omega>> \omega_0$ y hagamos el producto, **modulando** así cos ωt con f(t) :

$$f(t)\cos\omega t = A_0 \cos\omega t +$$

$$+\frac{1}{2}A_1\cos[(\omega+\omega_0)t+\phi_1]+\frac{1}{2}A_1\cos[(\omega-\omega_0)t-\phi_1]+$$

$$+\frac{1}{2}A_2\cos[(\omega+2\omega_0)t+\phi_2]+\frac{1}{2}A_2\cos[(\omega-2\omega_0)t-\phi_2]+$$

$$+\frac{1}{2}A_3\cos[(\omega+3\omega_0)t+\phi_3]+\frac{1}{2}A_3\cos[(\omega-3\omega_0)t-\phi_3]+\ldots$$

Aparecen las siguientes componentes:

a) Una componente de pulsación ω, si hay componente continua de f(t).

b) Un espectro completo de f(t) dividido por dos, desplazado a partir de la pulsación ω.

c) Un espectro simétrico del anterior con respecto a ω.

Diremos que el espectro de f(t) se ha **desdoblado** alrededor de ω igual que ocurrió en el párrafo anterior.

Si, por ejemplo, tomemos como f(t) la función rectangular periódica, obtenemos el resultado de la figura 2-10.

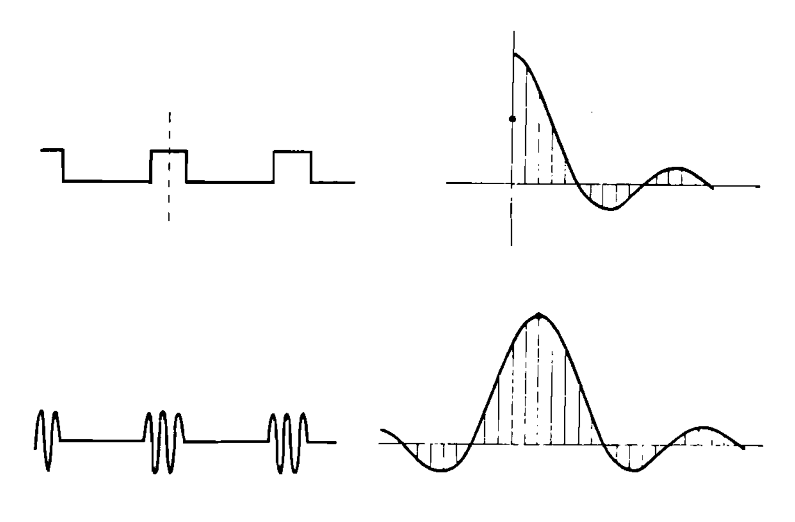


Fig.2-10

En el caso de que $\omega \ll \omega_0$ cada componente del espectro de f(t) da lugar a un desdoblamiento de valor ω por encima y por debajo. Aparece un espectro desdoblado y como desenfocado.

2.14.- Desarrollo exponencial de Fourier.

La familia de funciones complejas $e^{n\omega tj}$ es también un conjunto ortogonal completo, por lo que es posible descomponer la función f(t) en términos de la forma $C_n\ e^{n\omega tj}$, variando n entre menos y más infinito, lo que se llama **desarrollo exponencial de Fourier** o **desarrollo complejo** de la función f(t).

El desarrollo exponencial puede obtenerse fácilmente a partir del desarrollo descrito hasta ahora de la siguiente forma:

$$a_n \cos n\omega_0 t + b_n \text{ sen } n\omega_0 t =$$

$$\frac{a_n}{2}\left(e^{jn\omega_0 t}+e^{-jn\omega_0 t}\right)+\frac{b_n}{2j}\left(e^{jn\omega_0 t}-e^{-jn\omega_0 t}\right)=\frac{a_n-jb_n}{2}e^{jn\omega_0 t}+\frac{a_n+jb_n}{2}e^{-jn\omega_0 t}$$

es decir:

$$C_n=\frac{a_n-jb_n}{2}\ ;\ C_{-n}=\frac{a_n+jb_n}{2}$$

y en sentido inverso (sumando y restando las fórmulas anteriores):

$$a_n = C_n + C_{-n} \qquad\qquad b_n = j(\ C_n - C_{-n})$$

La componente continua que corresponde a n=0 es la misma en ambos desarrollos, $C_0 = a_0 = A_0$.

Los coeficientes a_n y b_n son números reales. De estas fórmulas se deduce que C_n y C_n son **complejos conjugados**. Su módulo es igual y vale:

$$|C_n|=\frac{\sqrt{a_n^2+b_n^2}}{2}=\frac{A_n}{2}$$

recordando el valor de A_n del desarrollo en cosenos. El argumento de C_n será :

$$\arg C_n = \text{arc tg } (\ -\ b_n\ /\ a_n\ ) = \phi_n$$

es decir, la fase del término n-simo del espectro.

Podemos ya hacer un resumen de conclusiones sobre el desarrollo de Fourier de la función f(t), que puede tomar una de las tres formas siguientes:

a) Trigonométrico en senos y cosenos con fase cero. No lo usaremos en este libro.

$$f(t) = a_0 + a_1 \cos \omega t + a_2 \cos 2\omega t + \ldots$$
$$b_1 \operatorname{sen} \omega t + b_2 \operatorname{sen} 2\omega t + \ldots$$

b) Trigonométrico en cosenos con fases distintas de cero. Es el que representa al **espectro real** de la función.

$$f(t) = A_0 + A_1 \cos (\omega t + \phi_1) + A_2 \cos (\omega t + \phi_2) + \ldots$$

c) Exponencial o complejo. Da lugar al **espectro complejo.**

$$f(t) = C_0 + C_1 e^{j\omega t} + C_2 e^{2j\omega t} + \ldots$$
$$C_{-1} e^{-j\omega t} + C_{-2} e^{-2j\omega t} + \ldots$$

donde los coeficientes y también los términos son complejos conjugados. Cada par de términos n-simos sumados dan un número real que es el término n-simo del espectro real.

El paso de uno a otro desarrollo se hace con las fórmulas halladas en este párrafo.

Los coeficientes C_n del desarrollo complejo pueden obtenerse directamente por la fórmula :

$$C_n = \frac{1}{T} \int_{t_0}^{t_0+T} f(t)\, e^{-jn\omega_0 t} dt$$

que se obtiene del mismo modo que se hizo la del desarrollo trigonométrico, es decir, escribiendo el desarrollo, multiplicando los dos miembros por $e^{-jn\omega t}$ e integrando a través de un período, con lo que resultan nulos todos los términos menos el enésimo.

2.15.- Espectro complejo de Fourier.

Si representamos los valores del módulo de C_n en función de la pulsación (o de la frecuencia) obtenemos un espectro en amplitud que es el espectro real **desdoblado**. Es decir, las amplitudes de C_n son la mitad de las A_n y además aparecen las de C_{-n} simétricas respecto al eje vertical de las de C_n . La fase de C_n es la misma que la de A_n y la de C_{-n} es la fase de C_n cambiada de signo. El espectro de fase se desdobla pues de otra forma que el de amplitud: El espectro para ω positivo es el mismo que el de fase real y el espectro para ω negativo es el simétrico respecto al origen (simetría respecto a un punto, con cambio de signo).

Resumiendo, podemos decir respecto al espectro complejo:

a) Aparecen términos de pulsación negativa, que no tienen ningún significado físico. Cada término complejo de pulsación

negativa debe sumarse en la realidad con el término complejo conjugado de pulsación positiva correspondiente, dando una suma real.

b) El espectro en amplitud se obtiene del espectro en amplitud real dividiendo las amplitudes por dos y tomando amplitudes simétricas respecto al eje de ordenadas (desdoblando).

c) El espectro de fase es el mismo espectro real de fase para pulsaciones positivas, y el simétrico respecto al origen para las negativas.

La figura 2-11 nos muestra el espectro real y el complejo para la función de impulsos rectangulares periódicos, procedente de la figura 2-8.

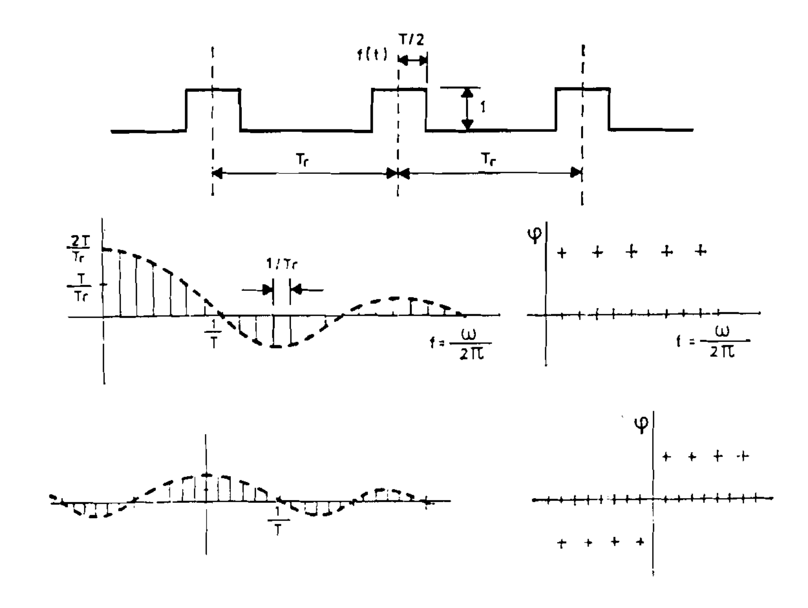


Fig. 2-11

El espectro complejo tiene la ventaja sobre el real de que se expresa con una fórmula sencilla y se utiliza por ello matemáticamente con facilidad. Por supuesto, el espectro obtenido con instrumentación es el real, pero también pueden representar el complejo, sin más que realizar las manipulaciones descritas en este párrafo.

2.16.- Espectro de funciones no periódicas.

Ya hemos dicho con anterioridad que **cualquier** función es equivalente a su desarrollo de Fourier en el intervalo elegido, independientemente de que sea periódica o no. Fuera del intervalo elegido como "período" el desarrollo de Fourier permanece periódico, mientras que la función puede tomar cualquier valor. Ya dijimos que esta sustitución es poco práctica porque

normalmente nunca puede decirse que los valores de la función fuera del intervalo no influyan en otros puntos del sistema, siendo entonces inexacta la sustitución.

Sin embargo, es posible desarrollar en serie de Fourier una función no periódica si elegimos un intervalo tan grande que estamos seguros de que los valores fuera de ese intervalo ya no van a influir en ningún sitio. Un modo poco riguroso pero expresivo de decirlo es afirmar que cualquier función no periódica es una función periódica de período infinito.

Vamos a hallar un espectro de este tipo con un ejemplo de gran aplicación práctica, el **impulso rectangular único** de altura unidad y ancho T. Podemos empezar con una sucesión de impulsos rectangulares periódicos como la representada en la figura 2-8, uno de cuyos impulsos es nuestro impulso rectangular "único" y podemos alejar los otros impulsos de la sucesión gradualmente, hasta conseguir un período muy grande. Podemos pues hacer crecer T ilimitadamente en la sucesión de impulsos periódicos, hasta convertirla en un solo impulso aislado. A veces se le llama a este impulso rectangular único " función puerta".

Observemos lo que ocurre en el espectro de la figura 2-8 cuando se hace crecer el período, que aquí es T_r .

a) Se conserva la forma general del espectro, que es la función de muestreo con nulos en las frecuencias múltiplos de 1/T. Las amplitudes, sin embargo, disminuyen de forma inversamente proporcional a T_r hasta anularse en el límite.

b) Las rayas espectrales se aproximan, el espectro se hace más denso. La distancia entre las rayas espectrales, que es $1/T_r$ tiende a cero al crecer el período, el espectro se hace **continuo.**

Definiremos un nuevo espectro igual al anterior multiplicado por el período T_r. Este nuevo espectro ya no varía nada en forma al crecer el período T_r , y lo único que varía es la distancia entre los puntos que se va acortando hasta hacerse una línea continua. Esta nueva definición de espectro es la que va a dar lugar a una nueva transformación, que sustituye al desarrollo de Fourier y que se llama **transformación de Fourier** .

2.17.- Transformación de Fourier.

Vamos a generalizar la operación del párrafo precedente. Para ello tomemos la fórmula que nos da los términos del espectro complejo:

$$C_n=\frac{1}{T}\int_{t_0}^{t_0+T} f(t)\,e^{-jn\omega_0 t}dt$$

hagamos el cambio de variable $n\omega_0 = \omega$,

$$C(\omega)=\frac{1}{T}\int_{t_0}^{t_0+T} f(t)\,e^{-j\omega t}dt$$

La función:

$$F(\omega)=\lim_{T\to\infty}[TC(\omega)]=\int_{-\infty}^{+\infty} f(t)\,e^{-j\omega t}dt$$

se llama **transformada de Fourier de f(t)** y representa el espectro de la función f(t) multiplicado por T y haciendo T infinito.

Del mismo modo podemos escribir el valor de f(t) como

$$f(t)=\sum_{-\infty}^{+\infty} C_n^{j\omega_0 nt}=\sum_{-\infty}^{+\infty} C(\omega)\,e^{j\omega t}$$

$$f(t)=\frac{1}{T}\sum_{-\infty}^{+\infty} TC(\omega)\,e^{j\omega t}$$

Como la distancia entre rayas espectrales vale ω_0, podemos llamar $\omega_0 = \Delta\ \omega = 2\pi/T$, con lo que

$$f(t)=\frac{1}{2\pi}\sum_{-\infty}^{+\infty} [TC(\omega)]\,e^{j\omega t}\Delta\omega$$

y pasando al límite para T infinito:

$$f(t)=\frac{1}{2\pi}\int_{-\infty}^{+\infty} F(\omega)\,e^{j\omega t}d\omega$$

que es la **transformada inversa de Fourier**.

La transformada de Fourier es una ayuda valiosa porque nos da el espectro de las funciones no periódicas, en especial los "impulsos" de cualquier forma. La existencia de la transformada de Fourier está condicionada a que exista la integral que la define.

Si se toma como variable independiente la frecuencia en lugar de la pulsación, las fórmulas de la transformación de Fourier y su inversa son :

$$F(f)=\int_{-\infty}^{+\infty} f(t)\, e^{-2\pi fjt}dt$$

$$f(t)=\int_{-\infty}^{+\infty} F(f)\, e^{2\pi fjt}df$$

2.18.- Potencia y energía normalizadas.

Se llama **potencia normalizada** de la señal periódica f(t) al valor:

$$W_0=\frac{1}{T}\int_{t_0}^{t_0+T}[f(t)]^2dt$$

que es la potencia eléctrica media de una tensión o una intensidad f(t) conectadas a una resistencia unidad. Si sustituimos f(t) por su desarrollo en cosenos, al elevar al cuadrado aparecen productos cruzados de diferente frecuencia cuyo valor medio es cero y quedan solamente las integrales de los términos cuadráticos, que valen 1/2. O sea:

$$W_0=\frac{1}{T}\int_{t_0}^{t_0+T}[A_0+A_1\cos(\omega_0 t+\phi_1)+\ldots]^2dt=A_0^2+\frac{1}{2}\sum_{1}^{\infty}A_n^2$$

o si tomamos el desarrollo complejo, teniendo en cuenta que $A_n = 2C_n$:

$$W_0=C_0^2+\frac{1}{2}\sum_{1}^{\infty}4|C_n|^2=C_0^2+2\sum_{1}^{\infty}|C_n|^2=\sum_{-\infty}^{+\infty}|C_n|^2.$$

Estas fórmulas dan la potencia como suma de las potencias normalizadas de los términos del desarrollo.

En el caso de las funciones no periódicas, se puede definir una **energía normalizada** como:

Si es F(ω) la transformada de Fourier de la función f(t) será:

$$E_0=\int_{-\infty}^{+\infty}[f(t)]^2dt$$

$$f(t)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}F(\omega)\,e^{j\omega t}d\omega$$

y podemos calcular:

$$E_0=\frac{1}{2\pi}\int_{-\infty}^{+\infty}f(t)\,[\int_{-\infty}^{+\infty}F(\omega)\,e^{j\omega t}d\omega]\,dt$$

o, cambiando el orden de integración:

$$E_0=\frac{1}{2\pi}\int_{-\infty}^{+\infty}F(\omega)\,[\int_{-\infty}^{+\infty}f(t)\,e^{j\omega t}dt]\,d\omega=\frac{1}{2\pi}\int F(\omega)\,F(-\omega)\,d\omega$$

y, teniendo en cuenta que $F(\omega)\ F(-\omega) = F(\omega)^2$ (complejos conjugados):

$$E_0=\frac{1}{2\pi}\int_{-\infty}^{+\infty}[F(\omega)]^2d\omega$$

Utilizando la transformación de Fourier de f (y no de ω) desaparece el 2π y queda:

$$E_0=\int_{-\infty}^{+\infty}[F(f)]^2df$$

lo que nos permite decir que **la energía normalizada del espectro es la misma de la función del tiempo.**

2.19.- Cálculo operacional.

La transformación de Fourier establece una correspondencia biunívoca entre las funciones f(t) y $F(\omega)$. A las operaciones en el "plano" t le corresponden otras operaciones en el "plano" ω. Por ejemplo, si la función f(t) se multiplica por una constante C, la transformada $F(\omega)$ se multiplica también por C, lo que

quiere decir que la transformada de Cf(t) es CF(ω), lo que se deduce también de que la transformación es lineal. También a la suma de funciones le corresponde la suma de espectros, es decir que la transformada de $f_1(t) + f_2(t)$ es $F_1(\omega) + F_2(\omega)$.

A otras operaciones en el plano t le corresponden operaciones no tan evidentes en el plano ω. Por ejemplo, **si se deriva f(t), la transformada de f(t) se multiplica por jω.** Es decir que si es F(ω) la transformada de f(t), la transformada de df(t)/dt es jωF(ω). Esta propiedad se deduce fácilmente de la fórmula de transformación. También si se integra en el plano del tiempo, la transformada **se divide por jω.**

Esta importantísima propiedad que permite convertir las operaciones de derivar o integrar en el plano t en multiplicar o dividir por jω en el plano ω facilita enormemente estas operaciones y permite convertir las ecuaciones diferenciales lineales en ecuaciones algebraicas en el plano ω, dando soluciones inmediatas.

Este tipo de cálculo, que consiste en utilizar transformaciones de este tipo se llama **cálculo operacional.** El factor (jω en este caso) de estas operaciones se llama **operador.**

En este libro vamos a utilizar cuatro tipos de recursos de cálculo operacional :

a) Desarrollo en serie de Fourier de funciones periódicas. **Espectros de funciones periódicas.**

En efecto, el desarrollo complejo de la función f(t) es:

$$f(t) = \sum C_n e^{jn\omega_0 t}$$

y su derivada es :

$$\frac{df(t)}{dt} = \sum jn\omega_0 C_n e^{jn\omega_0 t}$$

y como $jn\omega_0 = j\omega$, puede verse que se cumple la condición de que derivar es simplemente multiplicar por jω. **El operador es jω.**

b) **Transformación de Fourier**

En esta transformación, como hemos visto, el operador que permite derivar es jω.

c) **Transformación de Laplace.**

d) **Transformación jω (Fasores).**

De las transformaciones c) y d) hablaremos en los párrafos siguientes.

Los cuatro métodos son operacionales y en todos ellos hay un operador de derivación. Los dos primeros métodos permiten descomponer la función f(t) en funciones simples lo que les da un sentido físico mayor.

El uso de cálculo operacional puede hacerse con ayuda de tablas. Por ejemplo, si se quiere resolver una ecuación diferencial usando la transformación de Fourier se harán los siguientes pasos:

Se transforma la ecuación al plano ω, teniendo en cuenta que las derivaciones son simples multiplicaciones por jω.

Se resuelve la ecuación algebraica resultante, obteniendo la función solución en el plano ω.

Se vuelve al plano del tiempo con ayuda de unas tablas.

2.20.- Transformación de Laplace.

La transformación de Laplace es una generalización de la de Fourier. A cada función f(t) definida entre cero e infinito (obsérvese que no se tienen en cuenta los valores anteriores al origen) le corresponde una función F(s) donde s es una variable compleja (a diferencia de ω que era real). Se ha perdido el significado de ω como pulsación, s no tiene ese significado, se le puede atribuir el significado de una onda e_{at} cos ωt de pulsación ω y divergencia a (figura 2-12), siendo s= a + jω.

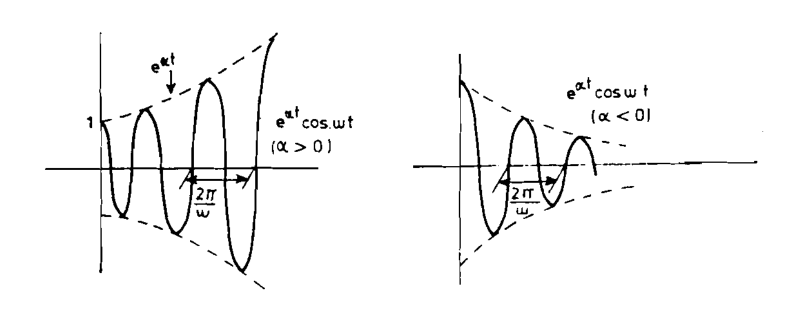


Fig. 2-12

La correspondencia se define por las fórmulas:

$$F(s) = \int_0^{\infty} e^{-st} f(t)\, dt$$

$$f(t) = \frac{1}{2\pi f} \int_{\sigma - j\infty}^{\sigma + j\infty} e^{st} F(s)\, ds$$

No es necesario que exista F(s) para todo valor de s, siendo suficiente que la integral de definición de F(s) converja para los valores de Re(s) > σ, siendo σ el "umbral" de s. Es decir, basta que exista F(s) para valores complejos a la derecha de la recta σ. Estas condiciones son más suaves que las de existencia de la transformada de Fourier F(ω). **La transformada de Fourier se obtiene directamente a partir de la de Laplace sustituyendo jω por s y viceversa.**

Al igual que pasaba con los desarrollos en serie, los valores de f(t) anteriores al origen no tienen relevancia matemática, pero en la realidad deben ser cero, evitando así que los resultados sean falsos por aparecer en otros puntos del sistema e influir en valores posteriores al origen.

La transformación es lineal. A la transformada de Laplace de f(t) se le llama también L f(t).

2.21.- Propiedad de derivación de la transformación de Laplace.

Trataremos ahora de hallar la transformada de df(t)/dt, conocida la de f(t). Sea:

$$\mathcal{L}f(t) = F(s) = \int_0^{\infty} f(t)\, e^{-st} dt$$

integrando por partes y teniendo en cuenta que e^{-st} es la derivada de $- e^{-st} /s$:

$$F(s) = [f(t)\,(-\frac{1}{s} e^{-st})]_0^{\infty} - \int_0^{\infty} (-\frac{1}{s} e^{-st})\, df(t)$$

como e^{-st} se anula para $t \to \infty$, será:

$$F(s) = \frac{f(0)}{s} + \frac{1}{s} \int_0^{\infty} e^{-st} \frac{df(t)}{dt} dt$$

de donde:

$$\int_0^\infty e^{-st}\frac{df(t)}{dt}dt=sF(s)-f(0).$$

El primer miembro es la transformada de la derivada de f(t). Es decir:

$$\mathcal{L}\frac{df(t)}{dt}=s\mathcal{L}f(t)-f(0)$$

Se obtiene así que la transformada de Laplace de la derivada se obtiene multiplicando por el **operador, s** y **restando el valor en el origen.**

El valor en el origen plantea un problema práctico en la determinación del valor de f(0), ya que la transformación de Laplace es especialmente adecuada para estudiar fenómenos discontinuos en el origen. Por ejemplo, es frecuente que la función f(t) sea el escalón unidad, es decir, una función que vale:

f(t) = 0 para $t < 0$
f(t) = 1 para $t > 0$
f(t) no definida en el origen.

La derivada de f(t) es la "impulsión unidad" o "función de Dirac" y se define como sigue:

f'(t) = 0 para $t < 0$
f'(t) = 0 para $t > 0$
f'(t) no definida en el origen

La integral de f(t)dt igual a 1 a través de cualquier intervalo que comprenda el origen.

Esta función se llama **impulsión unidad** y representa físicamente una percusión en el origen. También puede verse esta función como límite cuando $\delta \rightarrow 0$ del impulso rectangular de ancho δ y altura $1/\delta$. Este impulso sería la derivada de la función representada en la figura 2-13 en primer lugar, cuyo límite es el escalón unidad. En la parte inferior se representa el paso al límite de estas dos funciones.

En los sistemas técnicos, en los cuales se parte del reposo y aparecen perturbaciones que pueden ser discontinuas, es conveniente considerar que la perturbación está incluida en la integral de Laplace con todos sus efectos. Con ello puede garantizarse que todos los valores iniciales son cero y todos nuestros cálculos darán las soluciones a la derecha del origen, incluyendo posibles saltos o percusiones. La dificultad matemática de este punto singular se excluye fácilmente

considerando siempre estas funciones como casos límite de los de la figura 2-13. En el origen debe tomarse valor cero, es decir valor **a la izquierda** del origen.($f(0^-)$).

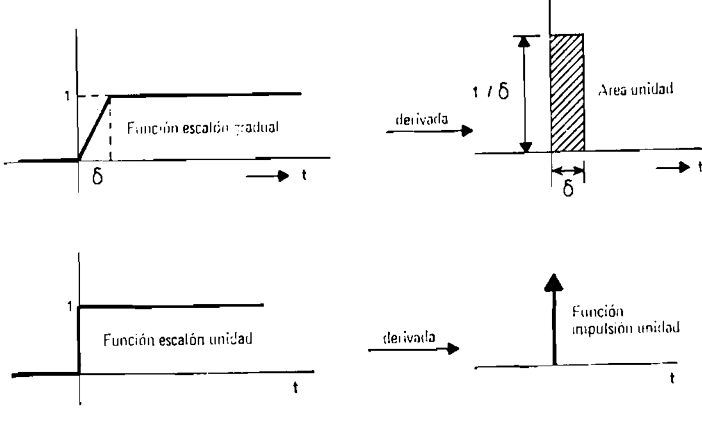


Fig. 2-13

Es preciso hacer notar que muchos libros, especialmente de matemáticas, hacen lo contrario, es decir, dejan fuera la discontinuidad y toman 1 como valor inicial. En ese caso es necesario calcular previamente todos los valores iniciales después del origen, integrando previamente a través del origen.

Por ejemplo, tomemos las dos funciones de la figura 2-14. La transformada de Laplace de cualquiera de las dos es 1/s, como es fácil calcular, sin que influya el valor en el origen. Pero al aplicar la regla de derivación resulta :

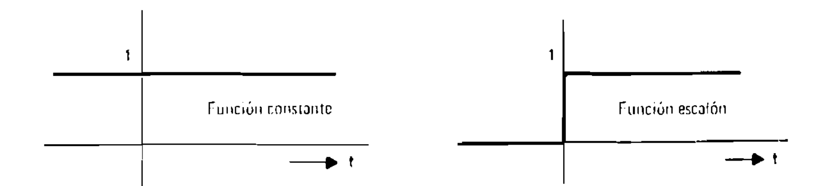


Fig. 2-14

Con nuestro sistema, la transformada de Laplace de la derivada será cero en el caso de la constante y será 1 en el caso del escalón unidad, lo que es correcto, ya que la derivada del escalón unidad debe ser la función impulsión unidad, de transformada 1. Si se toma el valor inicial a la derecha, resulta cero en ambos casos, con evidente confusión.

La propiedad de la derivación permite resolver ecuaciones diferenciales (en el plano del tiempo) transformándolas en algebraicas en el plano de Laplace, con la ventaja de que los valores iniciales se introducen fácilmente.

Sea, por ejemplo, la ecuación

$y'' + 3y' = x$ con valores iniciales y_0 e y'_0

Si llamamos Y a la transformada de Laplace de y, y X a la de X, teniendo en cuenta que:

$$\mathcal{L}y' = sY - y_0$$
$$\mathcal{L}y'' = s(sY - y_0) - y_0$$

Transformando los dos miembros de la ecuación al plano s, obtenemos:

$s^2 Y - sy_0 - y'_0 + 3sY - 3y_0 = X$

es decir:

$$Y = \frac{X}{s^2+3s} + \frac{(s+3)y_0 + y_0}{s^2+3s}$$

Los pasos a seguir serán :

a) Hallar en las tablas la transformada X de x, o **función imagen** de x.

b) Calcular Y a partir de X y de los valores iniciales con la fórmula anterior.

c) Hallar en las tablas la **función origen** cuya transformada es la calculada en b).

Por último haremos notar que si hacemos la hipótesis de que **partimos del reposo**, es decir, si suponemos que todos los valores iniciales (los de las variables y sus derivadas) son cero, el **operador** de Laplace es la variable **s**. Para derivar, se multiplica por s y para integrar se divide por s.

2.22.- Cambio de escala de tiempos.

Esta propiedad nos dice que si es F(s) la transformada de f(t), la transformada de f(at) , siendo a cualquier número real positivo, es :

$$\frac{1}{a} F[\frac{s}{a}]$$

Esta propiedad se deduce fácilmente de la fórmula de transformación y es también **válida para la transformación de Fourier**.

En el caso de la transformación de Fourier se comprende más fácilmente la interpretación física siguiente: Si se "contrae" la escala de tiempos, ($a < 1$) , es decir, se hace transcurrir el tiempo más deprisa, se "expande" el espectro, ya que tiempos más cortos equivalen a frecuencias más altas . El factor 1/a corresponde a una interpretación de "densidad" del espectro de Fourier .

2.23.- Teorema del retardo.

Si es F(s) la transformada de Fourier de f(t), siendo f(t)=0 para $t < 0$, la transformada de f(t - k) es e^{-sk} F(s). Este teorema del retardo se deduce también fácilmente de la fórmula de la transformación.

En el caso de la transformada de Fourier, el teorema del retardo es también válido sin restricciones, incluso para adelantos. No cambia el espectro de amplitud y sí el de fase.

2.24.- Teoremas del valor inicial y final.

La transformada de Laplace nos permite calcular los valores inicial y final f(0) y f(∞) a través de las fórmulas :

$$f(0)=\lim_{s\to\infty}sF(s)$$
$$f(\infty)=\lim_{s\to 0}sF(s)$$

Para demostrar estas fórmulas basta tener en cuenta que :

$$\mathfrak{L}\frac{df(t)}{dt}=sF(s)-f(0)=\int_0^\infty\frac{df(t)}{dt}e^{-st}dt$$

Haciendo tender $s \to 0$ se tiene :

$$sF(s)-f(0)=\int_0^\infty df(t)=f(\infty)-f(0)$$

de donde:

$$f(\infty)=\lim_{s\to 0}sF(s)$$

Del mismo modo, si se hace $s \to \infty$, resulta cero la integral (siempre que df(t) / dt no tenga puntos singulares) y se tiene:

$$f(0)=\lim_{s\to\infty} sF(s)$$

El hacer cero la integral obliga a eludir los puntos singulares. Por ello, el valor de f(0) que resulta por aplicación del teorema del valor inicial es **a la derecha del origen.**

2.25.- Representación de funciones sinusoidales por complejos.

La función A cos (ωt + ϕ) puede representarse de las formas siguientes:

a) Mediante un **vector** de módulo A y ángulo con el origen ϕ, llamado **fasor**.

b) Mediante un complejo de módulo A y argumento ϕ.

Ambas representaciones son equivalentes. La segunda, llamada **representación compleja**, es un instrumento importante de cálculo operacional.

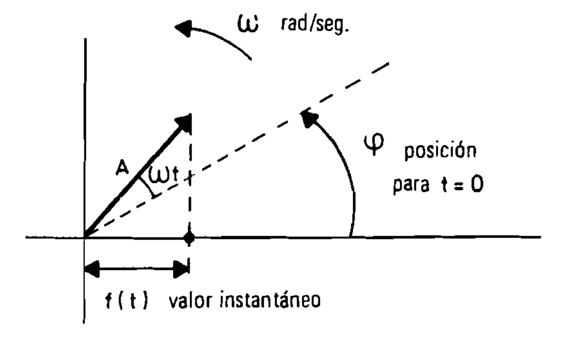
Se puede dar a la representación a), que luego conduce a la b), las interpretaciones físicas siguientes:

A cos (ωt + ϕ) es la proyección sobre el eje horizontal de un vector giratorio (ver figura 2-15) que comienza con el ángulo ϕy gira a ω rad./seg.

Si, posteriormente, se hace girar el papel en sentido contrario, se obtiene la figura de la derecha de 2-15 : El vector está quieto y la función del tiempo A cos (ωt + ϕ) es la proyección de este vector sobre un eje giratorio de velocidad ω en sentido contrario.

La interpretación vectorial (fasor) está ligada a la frecuencia ω. Con la última interpretación, un vector de pulsación ω_1 distinta no quedaría quieto en el "plano de fase", sino que giraría a la velocidad $\omega_1 - \omega$. Los vectores de la misma frecuencia sí se pueden representar en el mismo plano y se puede operar con ellos siguiendo las reglas del cálculo vectorial.

La representación compleja es una mera sustitución de los fasores por complejos convirtiendo el plano de coordenadas en un plano complejo. La tensión alterna f(t) = A cos (ωt + ϕ) queda pues representada por el complejo A cos ϕ+ j A sen ϕ, de módulo A y argumanto ϕ. También esta representación sólo está completa si se conoce ω o la frecuencia. Sólo entonces puede reproducirse la función del tiempo a partir del complejo.



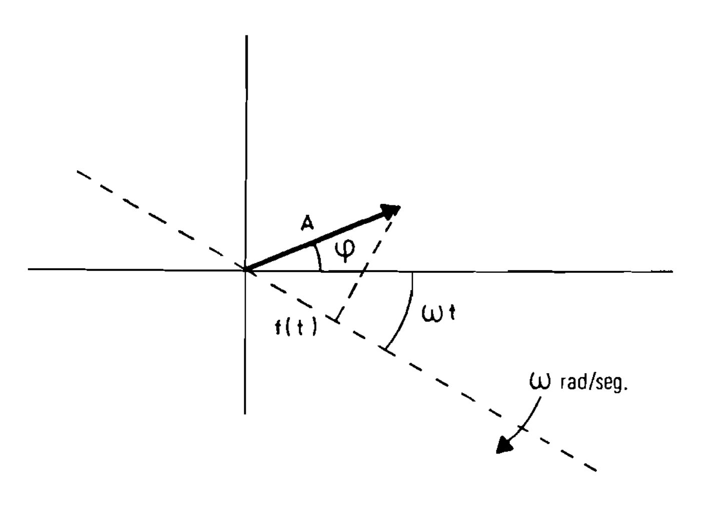


Fig. 2-15

2.26.- Regla de derivación.

La derivada de la función $f(t) = \cos(\omega t + \phi)$ es :

$$\frac{df(t)}{dt} = -\omega sen(\omega t+\phi) = \omega \cos(\omega t+\phi+\frac{\pi}{2})$$

lo que nos permite formular la siguiente regla para la derivación de funciones sinusoidales:

Para derivar una función sinusoidal (cosinusoidal) se multiplica por ω y se adelanta la fase $\pi/2$.

Si C es el complejo que representa a f(t), el complejo que representa a df(t) / dt se obtiene multiplicando por ω y adelantando $\pi/2$, es decir **multiplicando por $j\omega$. Se confirma así el carácter de operador de $j\omega$.**

2.27.- La transformación $j\omega$.

La representación compleja precedió bastante al uso del cálculo operacional, siendo realmente un recurso de cálculo operacional. Esto ha introducido un cierto desorden en la nomenclatura, por lo que es preciso señalar que la representación compleja es una verdadera transformación operacional válida para funciones sinusoidales y que el operador es $j\omega$, por lo que la llamaremos en adelante **transformación $j\omega$.** En la práctica es la más usada. Muchas veces se deja ω como "variable" y se representa la transformada **en función de ω**, lo que equivale a una representación gráfica de los espectros.

2.28.- Utilización del cálculo operacional.

Los sistemas electrónicos están caracterizados por relaciones entre señales. Por ejemplo, un amplificador está caracterizado por la relación entre la señal de salida y la de entrada.

Normalmente se definen las relaciones **suponiendo las señales sinusoidales** (cosinusoidales) y utilizando la transformación $j\omega$. Por ejemplo, diremos que la "impedancia" de una autoinducción es $j\omega L$, lo que quiere decir :

a) Que la relación entre los complejos que representan las señales de tensión entre sus extremos e intensidad que atraviesa la autoinducción es $j\omega L$, siendo L el valor de la autoinducción.

b) Por generalización de a), la relación entre el espectro complejo de la tensión y el de la intensidad es $j\omega L$. Es decir que la relación entre los espectros en amplitud es variable, igual a ω y que el espectro de fase de la intensidad está adelantado 90º. Lo mismo puede aplicarse a los espectros reales.

c) La relación entre las transformadas de Fourier de la tensión e intensidad será $j\omega L$.

d) Por analogía con c), teniendo en cuenta que la transformada de Laplace se obtiene a partir de la de Fourier sustituyendo $j\omega$ por s, la relación entre la tensión y la intensidad en el plano de Laplace será sL, es decir la **impedancia de Laplace de la autoinducción es sL.**

En la práctica es muy corriente la utilización directa del cálculo operacional sin especificar la transformación de que se trata e incluso sin especificar si se ha hecho una transformación. Por ejemplo, podemos decir que la entrada de un aparato es e y la salida s = 18e sin especificar el significado de las variables. En esa fórmula s y e pueden ser valores instantáneos s(t) y e(t) o bien las transformadas de Laplace o

Fourier de s(t) y e(t) o bien los complejos que representan ondas sinusoidales s y e. Esta claro que si escribimos s = 18je, s y e ya no pueden ser funciones del tiempo (**nunca pueden aparecer el tiempo y j en la misma fórmula**) pero sí cualesquiera de las otras transformadas.

No usaremos nunca en este libro recursos de cálculo operacional como la representación directa de A cosωt por $e^{j\omega t}$ o la separación posterior de parte real de la exponencial, que son lícitos, siempre que se hagan cuidadosamente las sustituciones correspondientes.

2.29.- Niveles de señal. Decibelios.

Las representaciones gráficas de las curvas que representan espectros o amplificaciones (relaciones de señales) pueden hacerse en escalas normales o en representaciones semilogarítmicas o logarítmicas.

Un caso particular de escala logarítmica de amplitudes lo constituye la medida en decibelios. Se llama **relación o diferencia de nivel en decibelios, (dB)** a la expresión :

$$20\log_{10}\frac{A}{B}$$

siendo A y B dos valores de la amplitud representativa de un fenómeno, en particular, tensión o intensidad, también llamados **niveles,** o bien:

$$10\log_{10}\frac{W_A}{W_B}$$

si se trata de dos **niveles** de potencia.Como la potencia eléctrica varía con el cuadrado de la amplitud de la tensión o intensidad (para una resistencia determinada) las dos expresiones dan el mismo resultado siendo W_A, W_B las potencias de las señales de amplitud A y B.

La relación en decibelios es una medida logarítmica. Al multiplicar la relación por un factor K, el valor en decibelios "aumenta" 20 log.K, ya que:

$$20\log\left[\frac{A}{B}K\right]=20\log\frac{A}{B}+20\log K$$

lo que justifica hablar de "niveles", que se suman en lugar de multiplicarse. El factor 20 es totalmente arbitrario y se usa para manejar números más cómodos. Al medir la relación en decibelios no es necesario especificar si se trata de tensiones, intensidades o potencias, ya que el número en decibelios es el mismo y lo que es diferente es la fórmula de aplicación que tiene un factor 10 en lugar de 20 cuando se trata de potencias.

El número de decibelios siempre está referido a una **relación**. Pero se puede medir en decibelios magnitudes absolutas tomando la relación con una unidad de referencia, que es el "nivel cero". Entonces se añade una letra que indica la unidad usada, por ejemplo, una W si ésta es un vatio, una m si es un milivatio, etc... y se habla de valores en dBW o dBm. Por ejemplo, se puede tomar como referencia una señal de 1 milivatio (o, incluso, especificar "1 milivatio sobre 600 ohmios). A partir de ahí, un nivel de 3 dBm quiere decir que la potencia es 2 milivatios, ya que 10 log 2 = 3 aproximadamente.

En la práctica se usan mucho los "desniveles" de 3 dB, equivalente a relación de potencia 2, 6 dB, equivalente a relación de amplitud 2 (o de potencia 4) y sus inversos, es decir, - 3 dB (relación de amplitud 1/2), - 6 dB (relación de potencia 1/2), así como las potencias de 10 en amplitud, representadas por 20, 40, 60, ... o -20, -40, -60, ...dB.

2.30.- La relación señal/ruido.

La señal es una variable, función del tiempo, que contiene información, generalmente tensión eléctrica, o sea, diferencia de potencial respecto a un origen de referencia. Junto con la señal, llegan otras variaciones de la variable de medida que **no** contienen información y perturban la detección y medida de la auténtica señal. Estas funciones no deseadas se llaman **ruidos**. Algunos ruidos llegan del exterior junto con la señal y otros se añaden durante la manipulación eléctrica de la señal.

Las fuentes de ruido mas importante son:

a) Interferencias o ruido producido por otras señales ajenas a la nuestra. En general se procura evitarlas cuidando el emplazamiento del equipo propio, limitando las potencias de las instalaciones cercanas, etc...

b) Ruido producido por alteraciones de componentes dentro del sistema. También se le llama "ruido de modulación" y tiene un espectro de bajas frecuencias. Se evita o disminuye mejorando la calidad del sistema.

c) Ruido termodinámico, debido a la agitación térmica de las partículas. Se trata de un **ruido blanco**, que cubre todo el espectro de frecuencias con amplitud aleatoria de valor medio casi constante. Es el más importante e inevitable. En nuestros sistemas lo dividiremos en **ruido exterior**, captado a través de sensores y antenas y **ruido propio**, producido por el propio sistema receptor de señal.

Los dos primeros tipos de ruido son evitables en gran medida tomando las precauciones adecuadas, por lo que sólo nos ocuparemos del ruido termodinámico, con el que es necesario contar siempre.

La **calidad** de una señal va a ser valorada en cierta manera por la **relación señal/ ruido**, que va a dar, como veremos la medida de la posibilidad de detectarla y de usarla. Para comprender esta cualidad de la señal, pongamos el siguiente ejemplo:

Supongamos que un grupo de personas hablan entre sí de forma aleatoria, pero todos a la vez. El conjunto de lo que oye un observador ajeno al grupo es lo que equivale al ruido. Supongamos ahora que el observador quiere entender lo que dice una determinada persona del grupo, cuyo discurso, que constituye la señal, está mezclado con el ruido general del grupo. Lo importante para el observador no es que esa persona hable muy fuerte, sino que hable más fuerte que los demás, que su voz sobresalga sobre las del grupo. Lo importante es la **relación** entre la señal y el ruido de fondo.

La **relación señal/ruido** en nuestros sistemas define la calidad o limpieza de la señal. Las señales serán más fácilmente detectables si su nivel es mucho mayor que el del ruido. La relación señal ruido mínima detectable depende de la habilidad del observador. En general convenimos en que en nuestros sistemas la relación **señal mínima detectable a la salida es uno, (cero decibelios)**. También puede decirse que convenimos en que la **señal mínima detectable es igual al ruido.**

Sea un sistema como el de la figura 2-16. A la **entrada** aparecerá una **señal S** y un **ruido exterior** acompañante R_0.

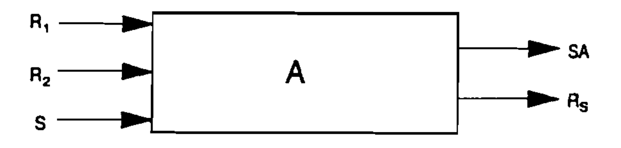


Fig. 2-16

El sistema genera un **ruido propio** que aparece a la salida con el sistema cortocircuitado a la entrada. El ruido propio se mide a la salida y se "refiere" a la entrada. Si el ruido medido a la salida es R_s y la amplificación del sistema es A (es decir, el sistema multiplica por A todo lo que entra en él), podemos suponer que existe un **ruido propio virtual R_i** a la entrada, que vale R_s / A y que, una vez amplificado, se convertiría en R_s. Siempre consideraremos este ruido propio (no existente) referido a la entrada. Conforme hemos definido anteriormente, **la señal mínima detectable a la entrada es R_i.**

La relación señal/ruido exterior (anterior al sistema) era S/R_0 y pasa a ser $S/(R_0 + R_1)$, inferior. A la salida, sigue teniendo este último valor, ya que tanto la señal como el ruido se multiplican por A. **El empeoramiento de la relación señal/ruido F** vale:

$$F=\frac{\frac{S}{R_0}}{\frac{S}{R_0+R_1}}=\frac{R_0+R_1}{R_0}\approx\frac{R_1}{R_0}$$

teniendo en cuenta que, generalmente, $R_1 >> R_0$. Esta relación suele darse en decibelios. La relación en decibelios tiene el mismo valor si las variables S y R están medidas en amplitud (por ejemplo en voltios eficaces) o en potencia (por ejemplo en vatios). **F, por lo tanto, además de ser el empeoramiento de la relación señal/ruido, es también la relación del ruido propio R_1 al ruido exterior R_0 y la relación señal/ruido mínima detectable R_1/R_0** que debe llegar al sistema para que la señal/ruido a la salida sea 1.

El valor **F** es una característica de calidad del sistema, y será tanto más bajo (recuérdese que es un empeoramiento) cuanto mejor sea el sistema. A este valor F se le llama **número de ruido** del sistema.

Por último digamos que el ruido exterior depende de la temperatura de agitación y, según puede demostrarse por razones termodinámicas, la potencia de ruido vale **$R_0 = K\ T\ \Delta f$**, siendo K la constante de Boltzmann, que vale 1,38 x 10^{-23} julios por grado Kelvin, T la temperatura absoluta en grados Kelvin (recuérdese que T= 273+t, midiendo t en grados centígrados) y Δf el ancho de banda en Hz. en el cual se reciben ruido y señal. Si el ruido se recoge a través de una resistencia de entrada R ohmios, se deduce fácilmente el valor de la tensión de ruido en voltios eficaces de la siguiente forma:

$$\frac{E_{R_0}^{\ 2}}{R}=KT\Delta f\ \text{, o sea ,}\ E_{R_0}=\sqrt{RKT\Delta f}$$

Teniendo en cuenta la definición de F, se calcula también fácilmente el valor del ruido propio a la entrada en voltios eficaces, que **coincide con la señal mínima detectable a la entrada**:

$$\frac{E_{R_1}^{\ 2}}{R}=FRKT\Delta f\ \text{o sea}\ E_{R_1}=\sqrt{FRKT\Delta f}$$

CAPITULO 3

REDES Y CIRCUITOS

3.1.- Introducción.

Los sistemas electrónicos transforman las señales por medio de **circuitos**, que son conjuntos de componentes conectados entre sí de forma ordenada para conseguir la transformación de las señales que llegan al circuito a su **entrada** y se envían a otro sitio, que puede ser otro circuito, una vez transformadas por la **salida** del circuito.

El elemento más simple que se utiliza para transformar señales es el o la **componente**. Un componente es un dispositivo básico, caracterizado por una ecuación, generalmente sencilla. Uniendo componentes entre sí se forma una **red**. Se llama red a cualquier conjunto de componentes conectadas entre sí. Una **red ordenada** para obtener un fin determinado se llama **circuito**. Un circuito es una red ordenada con una **entrada** y una **salida**, y realiza una determinada manipulación o transformación. Hay algunos circuitos sin entrada, cuando generan la señal en su interior, es decir, son generadores.

En este capítulo estudiaremos primero los componentes eléctricos más sencillos que no necesitan una fuente de energía eléctrica exterior para su funcionamiento y que se llaman **pasivos**. A continuación estudiaremos su combinación en forma de redes y circuitos. También estudiaremos algunas redes y circuitos con componentes activos o generadores. No entraremos en una definición exacta de redes pasivas y activas porque no tiene interés práctico y consideraremos redes pasivas a aquellas que no tienen generadores.

Debemos distinguir las redes o circuitos lineales de las que no lo son. En general , y mientras no se indique lo contrario, nos referiremos a circuitos lineales que son aquellos en que la salida cuando hay dos señales distintas es la suma de las salidas que corresponderían a cada una de las señales.

Damos por supuesto un conocimiento previo de Electricidad y de Redes Eléctricas y sus leyes fundamentales como lemas de Kirchoff, definiciones de impedancias, composición de

resistencias, etc... ya que nuestra exposición está encaminada al estudio de los **circuitos electrónicos** y sus peculiaridades.

3.2.- Componentes.

Los componentes son los elementos básicos con los que se construyen los circuitos y realizan las operaciones elementales de cálculo para la transformación de las señales. Los componentes están caracterizados por su ecuación, que relaciona las tensiones en sus terminales con las intensidades que circulan por ellos.Por ejemplo, en los componentes de dos terminales, la ecuación correspondiente relacionará la diferencia de potencial entre los terminales o tensión entre ellos con la intensidad que entrará por un terminal y que será igual a la que sale por el otro, por la condición de continuidad (la suma de intensidades que entran en un dispositivo debe ser cero ya que se supone que no se acumulan cargas eléctricas o, lo que es igual, que la capacidad respecto al resto del universo es cero).

A veces aparecen nuevos componentes. Para el ingeniero que las utiliza, basta con conocer sus ecuaciones características para encontrar enseguida aplicaciones concretas, ya que, realmente, no se dispone de una gran variedad de componentes. Por ejemplo, hasta la invención del elemento Hall no había ningún componente capaz de multiplicar dos variables.

A continuación se describen los componentes clásicos, insistiendo en las ecuaciones que los definen. Se describen solamente las propiedades que tienen relevancia electrónica y se da por supuesto un conocimiento previo de los componentes al nivel de Electricidad y Magnetismo.

3.3.- Resistencias.

Una resistencia es un componente constituido por un hilo arrollado sobre un tubo cerámico o una capa de depósito metálico cortada helicoidalmente o una capa de depósito de carbón. Su resistencia eléctrica (magnitud física que le caracteriza) debe ser lo más constante posible al variar la temperatura y otras condiciones ambientales o simplemente al sufrir pequeñas variaciones con el tiempo.

La ecuación de la resistencia es (figura 3-1):

$$E = IR \quad (\text{ ley de Ohm}).$$

donde E e I son los valores instantáneos de la tensión (diferencia de potencial entre los terminales) y la intensidad, ambos funciones del tiempo. El símbolo usado en la figura (una línea quebrada) corresponde a la simbología americana. En muchos libros y esquemas se usa la simbología DIN y la resistencia se representa por un rectángulo. En este libro seguiremos la norma americana que evita la confusión con otros elementos representados también por rectángulos.

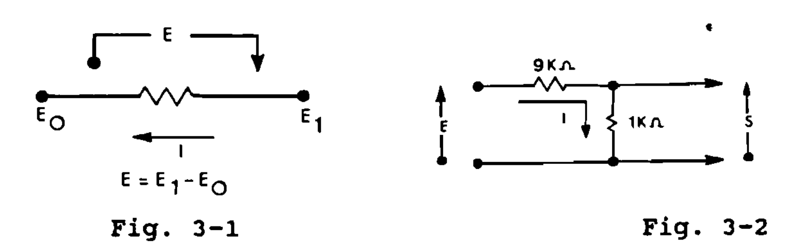


Fig. 3-1 Fig. 3-2

La resistencia efectúa una operación de cambio de variable (de tensión a intensidad o viceversa), multiplicando por una constante física, R, característica del componente. Algunas veces se llama al componente **resistor**, para diferenciarlo del valor de su magnitud característica, que es la **resistencia**. Sin embargo, en la práctica es corriente llamar resistencia a la componente, sin distinguirlo del valor de la magnitud resistencia.

Si se realiza la transformación de Fourier o la de Laplace de la ecuación de la resistencia se obtiene la misma ecuación que en el plano del tiempo. Por consiguiente, la ley de Ohm es una relación entre E e I que pueden ser.

a) Funciones del tiempo, E(t) e I(t).

b) Transformadas de Fourier E(w) e I(w) de E(t) e I(t).

c) Transformadas de Laplace E(s) e I(s) de E(t) e I(t).

Las resistencias se agrupan con fórmulas sencillas. Si están en serie **se suman**, ya que la intensidad es común y las tensiones se suman. Si están en paralelo, se suman los inversos, ya que la tensión es aquí común y las intensidades se suman. Esta propiedad es la "dual" de la anterior. En particular, el valor de una resistencia compuesta por R_1 y R_2 en paralelo es $R_1\ R_2/(R_1+R_2)$. Estas propiedades son bien conocidas en Electricidad.

En la resistencia se disipa una potencia $W = E^2/R = I^2R$ por efecto Joule. Esta potencia no debe sobrepasar un límite determinado para evitar un calentamiento excesivo. Comercialmente, se fabrican resistencias con límite de potencia de 1/2, 1, 2... W. Los valores comerciales también están normalizados, así como las tolerancias de fabricación, todo ello marcado generalmente con un código de colores.

3.4.- Condensadores

El condensador es un componente que utiliza una variable física auxiliar que es la cantidad de electricidad Q, con la que se relaciona la integral de la variable intensidad. En un condensador elemental, formado por dos placas separadas por un dieléctrico, se verifica :

$$Q= CE$$

siendo Q la cantidad de electricidad en culombios almacenada en

cada placa y E la tensión o diferencia de potencial en voltios entre las placas (valores instantáneos). La cantidad de electricidad positiva almacenada en una placa es igual a la cantidad de electricidad negativa almacenada en la otra, es decir que la carga total es cero. La intensidad que entra en una placa vale I = dQ/dt amperios y es igual a la intensidad que sale por la otra, como se indica en la figura 3-3. Por eso, aunque las cargas que "entran" por un lado no son las mismas que las que "salen" por el otro, las intensidades son iguales y se puede decir que la intensidad dQ/dt "circula" por el condensador o que lo atraviesa, ya que así aparece desde el exterior.

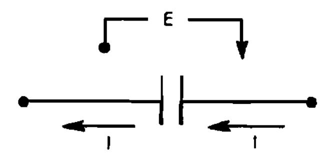


Fig. 3-3

La ecuación Q = CE es la definición de la **capacidad** C en **faradios**. Un faradio será un culombio partido por un voltio. El faradio es una unidad muy grande y se suelen usar los submúltiplos como el microfaradio, 1 μF = 10^{-6} faradios, el nanofaradio, 1 nF = 10^{-9} faradios y el picofaradio, 1 pF = 1 μμF= 10^{-12} faradios.

Derivando la ecuación Q = CE y teniendo en cuanta que I = dQ/dt se obtiene la ecuación del condensador:

$$I=C\frac{dE}{dt}$$

Esta ecuación define el comportamiento del condensador, caracterizado por su capacidad C. De esta ecuación se deducen enseguida algunas propiedades:

El condensador **integra** la intensidad que circula por él. Si hacemos que la intensidad sea una función determinada del tiempo, la tensión que aparece entre las placas será la integral de esa función, con un valor inicial que es la tensión o "carga" inicial. El condensador es un **integrador** automático.

La tensión del condensador **no puede cambiar bruscamente**, ya que esto exigiría una intensidad infinita. La intensidad sí puede hacerlo, puede ser discontinua, mientras que la tensión debe ser una función continua del tiempo.

El condensador **diferencia** la tensión que se le aplica. Si hacemos que esa función sea una función continua determinada del tiempo, la intensidad será la derivada de la función tensión. El condensador es un **diferenciador** automático.

En el caso de que E e I sean dos funciones sinusoidales del tiempo, podemos aplicar la regla de derivación de funciones sinusoidales, de que derivar es multiplicar por ω y adelantar $\pi/2$, o sumar $\pi/2$ a la fase, y de ahí deducir las siguientes relaciones entre E e I :

La intensidad que circula está **adelantada** respecto a la tensión $\pi/2$ y su amplitud es igual a la de la tensión multiplicada por ωC.

La tensión de un condensador está **retrasada** respecto a la intensidad $\pi/2$ y su amplitud es igual a la de la intensidad dividida por ωC.

Si aplicamos la transformación $j\omega$ a la ecuación del condensador, se obtiene :

$$I = j\omega CE$$

que se puede escribir:

$$\frac{E}{I}=\frac{1}{j\omega C}$$

El cociente de los complejos E e I en el plano complejo se llama **impedancia** y es una resistencia generalizada. La impedancia del condensador es :

$$Z_C=\frac{1}{j\omega C}$$

La impedancia del condensador para frecuencia cero (corriente continua) es infinito, o sea que para corriente continua la corriente **estacionaria** es cero, el condensador aparece como un circuito **abierto**, como si no existiese.

La impedancia del condensador es creciente con la frecuencia. Para frecuencias muy altas (tendiendo a infinito) tiende a cero, el condensador se comporta como un **cortocircuito.**

La aplicación del condensador como **integrador** se ilustra en la figura 3-4.

La figura 3-5 es un ejemplo de utilización del condensador para separar la componente continua de la alterna de una señal de entrada que contenga ambas componentes, basándose en que la impedancia del condensador es infinito para la componente continua y baja (eligiendo el valor de C suficientemente alto) para la componente alterna. Si consideramos sólo la componente continua, al transcurrir un tiempo suficiente el condensador estará cargado a la tensión E_0 y la tensión a la derecha sería 0. Para la componente alterna, despreciando la impedancia del condensador, la tensión a la derecha y a la izquierda serán

iguales. Puede decirse que el condensador no deja pasar la corriente continua y sí la alterna.

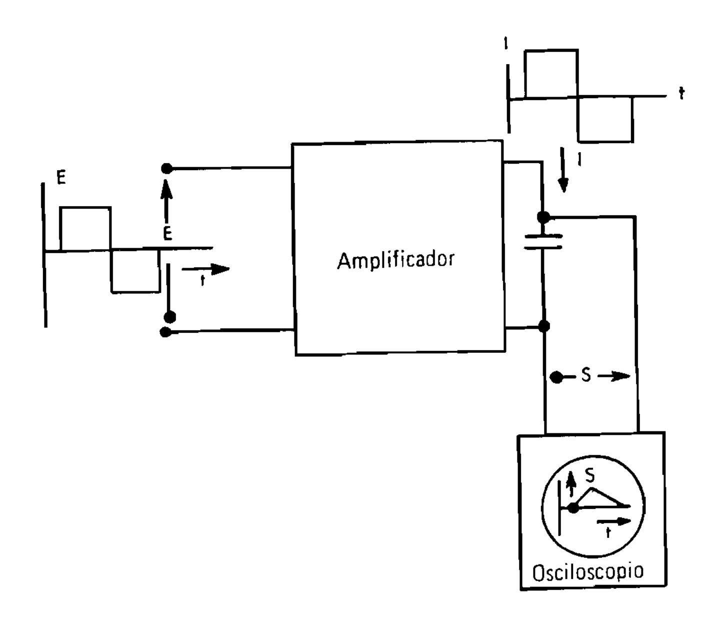


Fig. 3-4

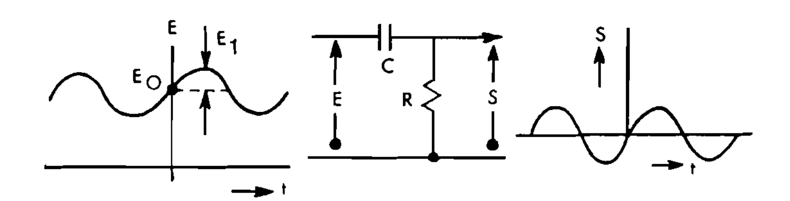


Fig. 3-5

Si aplicamos la transformada de Laplace a la ecuación del condensador, obtenemos I = sCE. Se puede también hablar de una impedancia de Laplace 1/sC, relación entre las transformadas de Laplace de la tensión y la intensidad.

En Electrónica existen diferentes tipos de condensadores según su construcción :

a) Condensadores de "aire", constituidos por placas metálicas separadas por dieléctrico de aire. El valor de la capacidad es relativamente bajo, por ejemplo, entre 100 y 600 pF. Se utilizan frecuentemente como condensadores variables o ajustables variando la posición relativa de las placas de un modo mecánico.

b) Condensadores cerámicos o de mica. Las placas metálicas están separadas por un dieléctrico como la mica u otro material

de alta constante dieléctrica. Las capacidades son más altas que en los de aire o el tamaño menor para la misma capacidad.

c) Condensadores de papel o papel metalizado. Están formados por largas tiras de papel o material plástico sobre el que se depositan finas capas de metal que constituyen los electrodos. Estas tiras se enrollan en forma de tubitos pequeños que pueden tener gran capacidad en poco volumen. Se obtienen así valores del orden de hasta unos microfaradios con componentes de unos centímetros.

d) Condensadores electrolíticos. Los electrodos son depósitos metálicos (electrolíticos) separados por un dieléctrico también depositado por métodos electroquímicos y todo ello sobre una base de material poroso, con lo que se obtiene la mayor superficie posible en un determinado volumen. Su capacidad es del orden de 1.000 veces superior a los anteriores.

3.5.- Inductancias.

La inductancia es un componente de **autoinducción L** . Como en el caso de los condensadores, se llama a veces **autoinducción** al componente mismo.

La inductancia está constituida por un **solenoide** o bobina con espiras que abrazan un flujo magnético común. La figura 3-6 representa el símbolo de la autoinducción y sus variables.

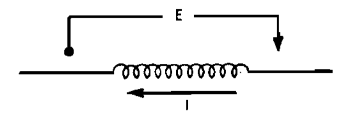


Fig. 3-6

La corriente que circula por el solenoide crea un flujo magnético longitudinal. Sus variaciones inducen una tensión proporcional a la derivada del cambio de flujo abrazado y, por lo tanto, proporcional a la derivada de la intensidad. Es decir, la ecuación de la autoinducción es :

$$E=L\frac{dI}{dt}$$

La unidad de autoinducción, definida a partir de la ecuación anterior, es el **Henrio**, que es la autoinducción de un solenoide si, aplicándola un voltio, deja pasar un amperio al cabo de un segundo. Esta unidad es también grande, por lo que, en la práctica se usan más los milihenrios (mHy) y los microhenrios (μHy).

Esta ecuación es la **dual** de la del condensador (cambiando tensión e intensidad) y las propiedades de la autoinducción son

las duales del condensador, es decir:

La intensidad es proporcional a la integral de la tensión aplicada, tomando como valor inicial el de la propia intensidad.

La intensidad no puede cambiar de forma discontinua.

Si se trata de funciones sinusoidales, se verifica también:

La tensión está adelantada $\pi/2$ respecto a la tensión y su amplitud es igual a la amplitud de la intensidad multiplicada por ωL.

La intensidad está retrasada $\pi/2$ respecto a la tensión y su amplitud es igual a la amplitud de la tensión dividida por ωL.

Utilizando la transformación $j\omega$, se obtiene la ecuación en el plano complejo:

$$E = j\omega L$$

De esta ecuación se deduce que la **impedancia de la autoinducción es $j\omega L$**. A partir de esta fórmula:

Para corriente continua, la autoinducción es un **cortocircuito**, su impedancia es cero.

La impedancia es creciente con la frecuencia. Para frecuencia infinito (o muy alta), la autoinducción es un circuito abierto, las frecuencias altas no pasan por la autoinducción.

La impedancia de la autoinducción en el plano de Laplace es **$s\omega$** .

Las aplicaciones de la autoinducción son las duales de las del condensador, si bien se aplica menos debido a la dificultad de obtener autoinducciones de calidad. También se aplica, junto con el condensador, para obtener efectos de resonancia.

Se utilizan tres clases de autoinducciones, según sea su **núcleo**, es decir el material por en el que aparece el flujo magnético:

a) De núcleo de aire (es decir, sin núcleo). Son muy exactas en su comportamiento, por no haber fenómenos de saturación ni histéresis. Se obtienen valores pequeños de autoinducción, de unos microhenrios.

b) De núcleo de hierro magnético.. Debido a la alta permeabilidad del hierro, se obtienen mayores valores de la autoinducción que con núcleo de aire, del orden de milihenrios. Son bobinas totalmente análogas a los transformadores de Electrotecnia. El inconveniente es la saturación y la histéresis del núcleo, que falsea en cierto modo la aplicación de la ecuación.

c) De núcleo de ferrita. La ferrita es óxido de hierro aglomerado con una resina sintética. Tiene propiedades de permeabilidad magnética parecidas a las del hierro con mejor comportamiento respecto a histéresis.

3.6.- Generadores.- Circuitos equivalentes.

Un generador es un componente activo constituido por un dispositivo, batería u oscilador, que genera una tensión continua entre sus bornes o una tensión determinada cualquiera, en la mayoría de los casos sinusoidal. Siempre que no se indique lo contrario nos referiremos con la denominación **generador** a los generadores sinusoidales, llamando simplemente baterías a los generadores de corriente continua.

El generador **ideal** es aquel cuya ecuación es simplemente **s = constante** independientemente de la intensidad suministrada, siendo s la tensión de salida, pudiendo ser una tensión continua (s(t) = constante) o alterna sinusoidal (complejo s constante). El símbolo correspondiente es el de la figura 3-7. En el caso del generador sinusoidal es necesario especificar la frecuencia o pulsación y la amplitud, pudiendo especificar ésta con su valor máximo o **de pico**, que generalmente se designa con el subíndice p o el valor **pico a pico**, doble del anterior,designado con el subíndice pp, o el valor **eficaz o efectivo**, designado con el subíndice ef, que es igual al de pico dividido por $\sqrt{2}$, que es el valor de cc. que da la misma potencia.

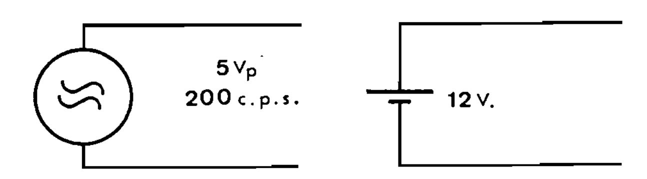


Fig. 3-7

El complejo s suele ser un número real, es decir de fase cero,lo que significa que esa tensión se toma como origen de fases, o sea su máximo como origen de tiempos.

También puede definirse un **generador de intensidad**, cuya ecuación será la misma s = constante, siendo s ahora la intensidad de salida.

En la práctica, no existen generadores ideales, teniendo todos una impedancia interna Z (compleja) de tal forma que la ecuación del generador pasa a ser :

$$s = e - Zi$$

siendo e la fuerza electromotriz e i la intensidad de salida ambas complejas. El generador ideal presenta una impedancia Z = 0 lo que quiere decir que la intensidad de salida no provoca una caída de tensión. En la figura 3-8 puede verse el circuito equivalente del generador real cargado con una impedancia

exterior Z_L. Este circuito consta de un generador ideal y la impedancia de salida aparente, que se dibuja en serie con la salida. Se llama **circuito equivalente** porque obedece a la misma ecuación que el generador real.

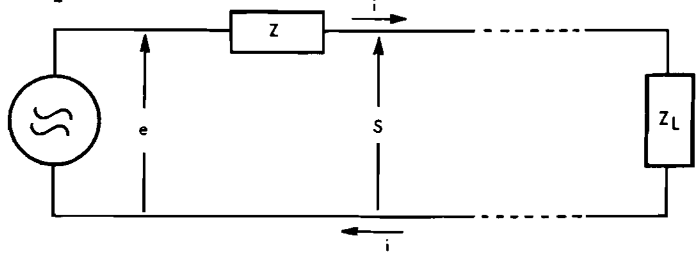


Fig. 3-8

Las ecuaciones que ligan las dos variables s e i son la del generador y la de carga, es decir:

s = e - Zi (ecuación del generador)
s = Z_L i (carga)

Resolviendo el sistema, se obtiene:

$$i=\frac{e}{Z+Z_L}$$
$$s=e\frac{Z_L}{Z+Z_L}$$

La tensión de salida vale s = e cuando Z es pequeña respecto a la carga, es decir $|Z| << |Z_L|$. Esta relación nos da un criterio para considerar como ideal a un generador real.(**impedancia interna despreciable frente a la carga**).

Cuando no se cumple la condición de que la impedancia interna sea despreciable frente a la carga, puede considerarse al generador real como un generador ideal al que se añade en serie una impedancia Z.

Cuando ocurre lo contrario (impedancia interna muy grande), la intensidad de salida es e/Z y se convierte en un generador de intensidad, por ser la intensidad independiente de la carga exterior.

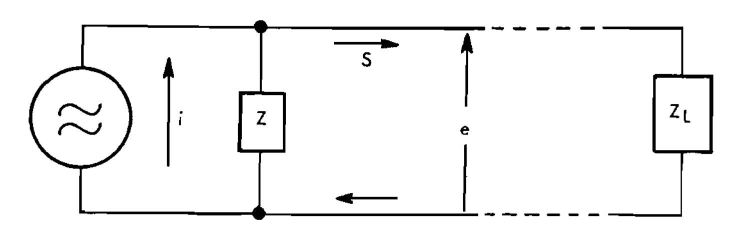


Fig. 3-9.

De forma análoga puede tratarse el problema de la impedancia interna de los generadores de intensidad. El generador real de intensidad se puede representar como se ve en la figura 3-9 por un generador ideal de intensidad i al que se agrega **en paralelo** una impedancia Z igual a la interna.

La ecuación del generador de intensidad es s = i - e/Z y si se carga con una impedancia exterior Z_L, la ecuación externa es e = Z_L s. Resolviendo esas dos ecuaciones se obtiene:

$$s = i\frac{Z}{Z+Z_L}$$

donde s se aproxima al valor i si $|Z| >> |Z_L|$.

En la práctica, los generadores de intensidad son poco frecuentes.

Cualquier generador de tensión puede "pasarse" a generador de intensidad observando que la ecuación del generador de tensión s = e - Zi puede escribirse de la forma i = e/Z - s/Z, que es la ecuación de un generador de intensidad e/Z con impedancia interna Z en paralelo. En la figura 3-10 pueden verse los circuitos que corresponden al generador de tensión y al equivalente de intensidad.

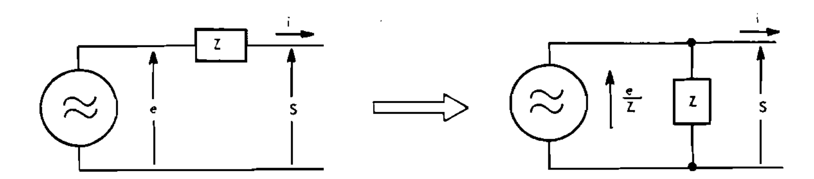


Fig. 3-10

Los dos circuitos de la figura 3-10 se llaman **equivalentes** porque obedecen a **la misma ecuación**.

Del mismo modo, el generador de intensidad de la figura 3-9 puede transformarse en el generador de tensión **equivalente** representado en la figura 3-11, simplemente escribiendo la ecuación s = i - e/Z en la forma e = Zi -Zs.

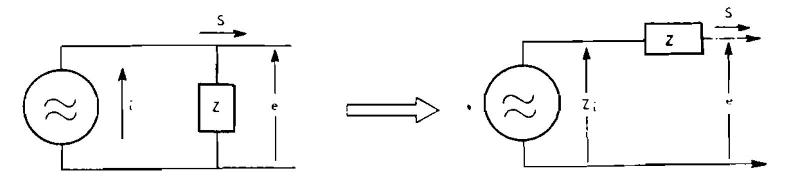


Fig. 3-11

En todos estos casos de **circuitos equivalentes**, no hay que olvidar que las ecuaciones relacionan variables **exteriores** y que la sustitución de un circuito por otro equivalente es válida a todos los efectos exteriores.

También es interesante notar aquí que la salida de un generador es cero a cualquier frecuencia que no sea la propia, lo que puede expresarse así :

Un generador ideal de tensión es un **cortocircuito** (tensión cero) a cualquier frecuencia que no sea la propia. En particular, una batería es un cortocircuito a cualquier frecuencia no nula y un generador de alterna es un cortocircuito para corriente continua.

Un generador ideal de intensidad es un **circuito abierto** (intensidad cero) a cualquier frecuencia que no sea la propia.

3.7.- Redes.

Los componentes se agrupan para cumplir su misión operativa formando **circuitos**. Un conjunto de componentes unidos entre sí se llama una **red**. Un circuito es una red con una función determinada y dispone de un o varias **entradas** y una o varias **salidas** de señales. Excepcionalmente puede tener sólo salidas, cuando desempeña la función de generar señales.

A continuación repasamos algunas de las propiedades generales de las redes, que suponemos previamente conocidas, con objeto de subrayar su aplicación a los **circuitos electrónicos**. Aunque no existe una diferencia esencial entre circuitos eléctricos y electrónicos, podríamos anotar que en los circuitos eléctricos en general suele manejarse una frecuencia única con poca o ninguna información (es decir, no se trata de señales) y, en cambio, tiene mucho interés la **potencia** o la energía an general, cuyo manejo es el objetivo final de esos circuitos. En los circuitos electrónicos, se manejan frecuencias distintas o incluso señales de frecuencia variable. La potencia suele ser muy pequeña y no tiene relevancia desde el punto de vista de la información.

Las redes o circuitos se representan por **esquemas** eléctricos que representan topológicamente los componentes y sus conexiones. Las uniones entre componentes están representadas por líneas, independientemente de que en la realidad tengan esa forma (cables o hilos) o la conexión esté hecha a través de un circuito impreso.

Todas las líneas que se unen en algún punto, hasta su llegada a los componentes, constituyen un **nudo**. El nudo está a un **potencial** eléctrico que le caracteriza, con una referencia de potencial común a toda la red, que es el nudo de potencial cero.

La suma de todas las intensidades que concurren en un punto es cero (primer lema de Kirchoff). Este es el principio de "continuidad" que expresa que el nudo no acumula cargas o que su

capacidad respecto al universo es cero.

Los componentes son las **ramas** de la red. Por ellas circulan **intensidades** que son las que concurren en los nudos. Los componentes ligan con sus ecuaciones las diferencias de potencial entre los nudos con las intensidades que circulan por las ramas.

El origen de potenciales es arbitrario. Se puede atribuir potencial cero a cualquier nudo de la red o bien al nudo que está conectado a **tierra**, que es lo que se suele hacer en la práctica. Esta tierra es ficticia, se trata sólo del chasis común a todo el equipo o, simplemente un nudo con muchas conexiones al que se ha llamado **tierra**. Cuando se conectan varias redes entre sí, el origen o tierra debe ser común, es decir, las diferentes tierras deben estar conectadas entre sí. Algunas veces la conexión exterior a tierra se hace realmente a la tierra real, garantizando así un origen de potenciales común con otros equipos. Fijando los potenciales en los nudos (en realidad son diferencias de potencial con tierra) como variables, se cumple automáticamente el "segundo lema de Kirchoff", que dice que la suma de las diferencias de potencial en las ramas que constituyen un lazo cerrado es cero.

La suma de las intensidades que entran en la red o circuito es también cero, es decir que suponemos que **toda** la red tiene capacidad cero respecto al exterior, no acumula cargas.

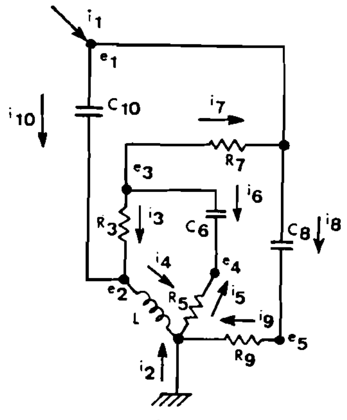


Fig. 3-12

Dos redes se consideran eléctricamente equivalentes si tienen los mismos componentes y coinciden los componentes que concurren en cada nudo. Se dice entonces que existe una correspondencia topológica de redes y nudos. Por ejemplo, las redes de las figuras 3-12 y 3-13 son equivalentes, se puede llevar una a la otra deformando la figura pero conservando la composición de ramas y nudos. En la red existe también un número de "polígonos" que está relacionado con el de redes y ramas por la fórmula $P + N - 1 = R$, siendo P el número de polígonos independientes, N el número de nudos y R el de ramas. Esta fórmula se deduce por consideraciones topológicas y no tiene interés para este libro porque no utilizaremos las relaciones de

polígonos para resolver las redes.

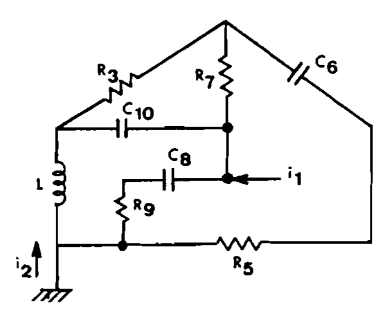


Fig. 3-13

Las **variables internas** de la red son las intensidades de las ramas y los **potenciales** de los nudos, excepto el de tierra, que es cero por definición. Por consiguiente, el **número de incógnitas** de la red será R + N - 1.

Además de las variables internas hay otras **de ligadura**. Si se conecta la red al exterior por un solo punto, por ejemplo la toma de tierra, por esa ligadura no circula ninguna corriente, ya que la ecuación de continuidad de la red (la no acumulación de cargas) obliga a que la intensidad de una sola ligadura sea cero. Si existe otra conexión, como en la figura 3-12, la suma de intensidades de las dos conexiones es cero, con lo que sólo existe una variable de ligadura, ya que $i_2 = - i_1$. En general, si hay n conexiones exteriores, sólo hay n - 1 **incógnitas de ligadura** o **intensidades de ligadura**, ya que la n-sima es consecuencia de las demás por tener que ser cero la suma de intensidades exteriores. Generalmente no se considera como ligadura la conexión de tierra.

Las **ecuaciones** de la red son las ecuaciones de los componentes de las ramas y la ecuación de continuidad de los nudos. La ecuación del último nudo (puede ser el de tierra) es consecuencia de las demás, ya que todas las intensidades suman cero, por lo que el **número de ecuaciones independientes** en una red aislada (sin conexiones exteriores) es R + N - 1.

El sistema de ecuaciones que define el comportamiento de la red es un sistema de R + N - 1 ecuaciones diferenciales con R + N - 1 incógnitas internas funciones del tiempo. En realidad, las ecuaciones de las ramas pueden ser diferenciales y las de los nudos no lo son. El "orden" del sistema corresponde con el número de componentes condensadores y autoinducciones cuyas ecuaciones son diferenciales de primer orden. Los valores iniciales del sistema corresponden a las cargas e intensidades iniciales por condensadores y autoinducciones respectivamente. La resolución en el momento cero del sistema permite convertir estos valores en valores iniciales de las variables.

Si no hay ligaduras exteriores, el sistema de R + N -1 ecuaciones con R + N - 1 incógnitas (todas "interiores" de la

red) está determinado y la red puede "resolverse", es decir, hallar el valor de todas sus variables en función del tiempo. Si hay n ligaduras exteriores (n + 1 conexiones), es decir, una intensidad exterior por cada conexión exterior a partir de la segunda, el sistema tiene n grados de libertad.

Hemos supuesto tácitamente que existía una componente por cada rama, pero es posible agrupar previamente componentes y formar ramas algo más complicadas, con ecuaciones obtenidas de la agrupación de componentes.

En el caso de tratarse de funciones sinusoidales, las ecuaciones diferenciales se reducen a normales, aplicando cálculo operacional antes o después de aplicarlas. El procedimiento más cómodo es aplicar **impedancias** complejas, lo que equivale a aplicar la transformación $j\omega$.

Por ejemplo, la red de la figura 3-12 tiene como incógnitas interiores los valores de e_1, e_2, e_3, e_4, e_5, i_3, i_4, i_5, i_6, i_7, i_8, i_9, i_{10}, 13 variables internas. Las ecuaciones son para ondas sinusoidales de pulsación ω son :

Ecuaciones de las ramas (8):

$$Ec.\ Ramas \begin{cases} e_3 - e_2 = R_3 i_3 \\ e_2 = j\omega L i_4 \\ -e_4 = R_5 i_5 \\ e_3 - e_4 = \dfrac{1}{j\omega C_6} i_6 \\ e_3 - e_1 = i_7 R_7 \\ e_1 - e_5 = \dfrac{1}{j\omega C_8} i_8 \\ e_5 = R_9 i_9 \\ e_1 - e_2 = \dfrac{1}{j\omega C_{10}} i_{10} \end{cases}$$

Ecuaciones de los nudos (5):

$$Ec.\ Nudos \begin{cases} i_1 + i_7 - i_{10} - i_8 = 0 \\ i_{10} + i_3 - i_4 = 0 \\ -i_3 - i_6 - i_7 = 0 \\ i_6 + i_5 = 0 \\ i_8 - i_9 = 0 \end{cases}$$

Existen 13 ecuaciones y 14 incógnitas incluida la variable exterior o de ligadura i_1, es decir hay un grado de libertad. Se puede resolver el sistema calculando todo en función de i_1, es decir considerando i_1 como dato.

En el caso más general de que no se trate de funciones

sinusoidales (es decir, que i_1 no sea sinusoidal), las ecuaciones que corresponden a las ramas de L, C_6, C_8 y C_{10} son

$$e_2=L\frac{di_4}{dt};\ i_8=C_8\frac{d(e_1-e_5)}{dt}$$
$$i_6=C_6\frac{d(e_3-e_4)}{dt};\ i_{10}=C_{10}\frac{d(e_1-e_2)}{dt}$$

El número de ecuaciones es el mismo de antes pero ahora hay que conocer cuatro valores iniciales correspondientes a las cargas iniciales de los tres condensadores y la intensidad inicial por la autoinducción y también $i_1(t)$ completa, es decir para todo valor de t. Se puede aplicar la transformada de Laplace, contando con los valores iniciales o, si se parte del reposo, la transformada con impedancias de Laplace.

3.8.- Principio de superposición.

Las ecuaciones de las redes tienen tres clases de términos:

a) Los que dependen de las variables "internas" de la red.

b) Los que dependen de las ligaduras o conexiones exteriores.

c) Los que dependen de los generadores.

En general las ligaduras aportan **ecuaciones de ligadura**, una por cada variable de ligadura, que, a su vez, pueden ser ecuaciones con o sin generadores. Las ligaduras pueden ser **datos**, que simplemente son constantes y pueden considerarse como generadores, o están ligadas con ecuaciones con generadores. Con ello, el sistema de ecuaciones queda establecido como un sistema **lineal** con términos independientes que son los generadores.

Si no hubiese generadores(ni ligaduras que pueden considerarse generadores) las ecuaciones serían **homogéneas**, y la única solución estacionaria es cero. Es lógico que si no se aplica algún generador, directamente o a través de ligaduras, no aparezcan señales en la red (excepto las que correspondan al régimen transitorio por valores iniciales). Supondremos en adelante que partimos del reposo o que sólo estudiamos el régimen estacionario. En ese caso, la única solución del sistema homogéneo es cero. La solución distinta de cero aparece cuando hay generadores o datos (ligaduras) que pueden considerarse como generadores, y que hacen que las ecuaciones dejen de ser homogéneas.

Las ecuaciones lineales permiten descomponer la solución del sistema con n generadores en suma de n soluciones, obtenida cada una de ellas con **un generador** y **anulando los demás**. Este es el **principio de superposición**. Como esta descomposición puede ser anterior a la aplicación de la transformación $j\omega$, las soluciones

parciales pueden ser a distintas frecuencias.

El principio de superposición se aplica en la práctica de la siguiente forma:

Se considera un primer generador de frecuencia f_1 o pulsación $\omega_1 = 2\pi f_1$ y se dejan los otros generadores de la misma frecuencia, si los hay Se **cortocircuitan** (que es lo mismo que anular) los demás generadores. A continuación se resuelve el sistema formado por las R + N -1 ecuaciones de la red más las de ligadura (que pueden ser del tipo de "variable de ligadura igual a constante" o más complicadas). Recuérdese que también las "constantes" de ligadura deben anularse si no corresponden a la misma frecuencia f_1.También pueden "introducirse" las ligaduras sustituyendo su valor en las ecuaciones de la red, con lo que quedan eliminadas como variables.

A continuación se procede igual con la segunda frecuencia f_2, anulando (cortocircuitando) todas las demás, incluida la primera.

Así se continúa con todas las frecuencias presentes en la red.

La solución para cada variable es la suma de las soluciones parciales **en el plano del tiempo**. Recuérdese que **no** se pueden sumar señales de distinta frecuencia en el plano complejo.

3.9.- Circuitos.

En la práctica de la Ingeniería Electrónica no se trabaja con redes desordenadas sino con circuitos funcionales. Los circuitos que sólo tienen componentes pasivos se llaman también pasivos, mientras que los que tienen componentes activos se llaman activos.

Los circuitos electrónicos tienen siempre una **salida**, ya que para que sean útiles hay que extraer señales. La salida puede estar constituida por dos **terminales**, o conexiones exteriores, de forma que la señal de salida es la tensión o diferencia de potencial entre ellos, o un terminal, y la salida es la señal que aparece entre este terminal y tierra (que también se conecta al exterior). También pueden tener varias salidas.

Los circuitos electrónicos pueden también tener una **entrada** de señal con uno o dos terminales (cuando es uno, la referencia es tierra). También pueden tener varias entradas.

Los circuitos activos necesitan una **alimentación** o fuente de energía. La conexión de alimentación **no tiene relación con las señales**. Generalmente la alimentación es una conexión a la que hay que suministrar energía con una fuente de corriente continua, que, aunque se le suele llamar así, es un generador de tensión continua con baja impedancia. Se representa como una **batería**, aunque muchas veces es un circuito **fuente de alimentación**, que transforma la tensión alterna de la red de suministro de energía

en una tensión continua apropiada, 12 V., por ejemplo.

3.10.- Circuitos de dos terminales.

Los circuitos de dos terminales son redes funcionales activas o pasivas que se unen al resto del sistema por dos conexiones. El circuito de dos terminales se representa como se ve en la figura 3-14 por un rectángulo o bloque con dos terminales, y su comportamiento está definido por dos variables, que son la tensión e entre los terminales y la intensidad i que sale por uno de ellos y **vuelve** por el otro. Existe una **única** intensidad porque la suma de las intensidades que entran en el circuito debe ser cero, o, lo que es lo mismo, la que sale por un terminal igual a la que entra por el otro.

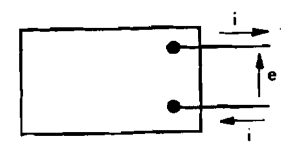


Fig. 3-14

Los dos terminales son dos nudos de la red. La variable e es interna y es igual a la diferencia de potenciales de esos dos nudos. La variable i es una única ligadura exterior.

En la práctica los circuitos de dos terminales representan las **cargas** si no tienen generadores, o los **orígenes de señal** cuando contienen generadores.

Si el circuito tiene R ramas y N nudos, hay R + N -1 variables internas más una externa, es decir R + N variables en total y R + N - 1 ecuaciones del circuito (internas). Existe un grado de libertad. Eliminando todas las variables menos e, i, es decir R + N - 2 variables internas, necesitaremos R + N - 2 ecuaciones y quedará **una ecuación lineal entre e , i.**

Si suponemos señales sinusoidales y usamos transformación compleja, la ecuación **de la red** podrá siempre escribirse de la forma :

$$e = A - Bi$$

siendo A y B complejos funciones de la frecuencia.

Si la red no tiene generadores, A = 0 porque la ecuación debe ser homogénea, y queda sólo e = - Bi, que nos dice que **el circuito de dos terminales sin generadores es equivalente a una impedancia**. El signo - de la ecuación procede de la designación del signo de i saliendo del circuito, como corresponde a un generador y no a una impedancia. El circuito equivalente está representado en la figura 3-15.

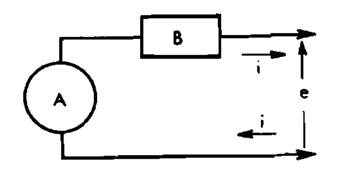


Fig. 3-15

3.11.- Teorema de Thevenin.

La interpretación física de la ecuación e = A - Bi nos da un método sencillo de hallar A y B por simple observación del circuito. Esta ecuación es la de un generador de fuerza electromotriz e y impedancia interna B, lo que constituye el **circuito equivalente**, de la figura 3-15.

El significado de las dos constantes es el siguiente: Si A = 0, B es la relación e / (-i) , es decir es la relación tensión / intensidad de entrada (-i es la intensidad de entrada, vista desde el exterior), o, lo que es lo mismo **la impedancia que se ve desde el exterior**, con A = 0, o sea anulando o **cortocircuitando todos los generadores**.

A es la tensión de salida sin intensidad, con i = 0, es decir, **sin carga**.

Podemos así formular completo el teorema de Thevenin:

Un circuito de dos terminales es equivalente a un generador cuya fuerza electromotriz es la que aparece entre los terminales sin carga y cuya impedancia interna es la que se ve desde fuera cortocircuitando los generadores.

El teorema de Thevenin permite la sustitución de circuitos muy complicados por otros **equivalentes** (con las mismas ecuaciones) más sencillos. Por ejemplo, apliquemos el teorema de Thevenin al **divisor de tensión** de la figura 3-16, que es un divisor 10:1 con el generador incluido.

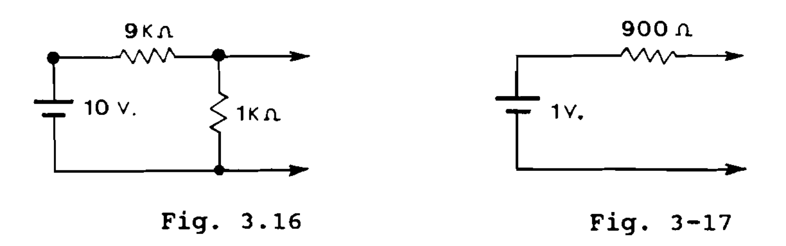


Fig. 3.16 Fig. 3-17

El valor de A se deduce fácilmente de la red interna, calculando la intensidad que pasa por las dos resistencias (sin carga exterior) que vale 10 V./ 10 KΩ = 1 mA. La caída de tensión en la resistencia de 1KΩ es A = 1 V. Luego, A = 1 V.

El valor de B se obtiene cortocircuitando la batería de 10 V. con lo que se "ven" las dos resistencias en paralelo y valen:

$$\frac{9.000 \times 1.000}{10.000} = 900\,ohmios$$

El circuito de la figura 3-17 es el equivalente de Thevenin del divisor de la izquierda. A la vista de este resultado se aprecia enseguida que el divisor de tensión es un buen reductor de la tensión siempre que no se cargue apreciablemente, ya que la resistencia de salida es demasiado alta si se eligen altos los valores de las resistencias. Si se eligen valores bajos, la resistencia de salida sería baja, pero se pierde mucha potencia en el propio divisor.

Si la ecuación del circuito equivalente se escribe en forma de generador de intensidad

$$i = A - \frac{e}{B}$$

se obtiene el teorema dual de el de Thevenin, que se llama teorema de **Norton** y que dice así :

Una red de dos terminales es equivalente a un generador de intensidad de valor el que circularía entre los terminales cortocircuitándolos (A = i si e = 0) y de impedancia interna la que se ve desde fuera anulando todos los generadores (cortocircuitando los de tensión y abriendo los de intensidad). El teorema de Norton se aplica muy pocas veces en Electrónica.

3.12.- Cuadripolos.

Un tipo de circuitos de gran aplicación electrónica es el circuito de cuatro terminales, dos de **entrada** y dos de **salida**, llamado **cuadripolo**. Como caso particular, están incluidos los circuitos de tres terminales en que uno de ellos (generalmente el de tierra) es común a la entrada y a la salida, y que puede estudiarse formalmente como de cuatro terminales. La figura 3-18 representa ejemplos de redes de cuatro terminales.

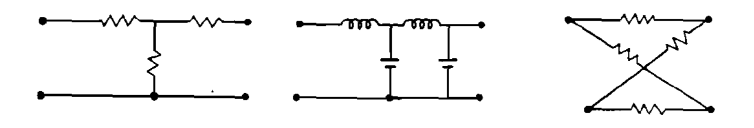


Fig. 3-18

El cuadripolo se representa como se ve en la figura 3-19 en forma de un rectángulo con entrada y salida. Supondremos señales sinusoidales y utilizaremos la transformación $j\omega$.

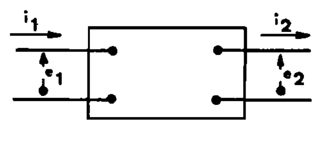


Fig. 3-19

El cuadripolo representa la transformación de las señales entre la entrada y la salida. Normalmente va precedido de un generador y va seguido de una carga, aunque puede haber otros cuadripolos intermedios. Describiremos las ecuaciones y parámetros suponiendo que se trata de cuadripolos **pasivos**, aunque algunas propiedades se generalizan fácilmente en casos particulares que veremos más adelante.

Las variables de entrada son la tensión de entrada y la intensidad de entrada. Sólo hay una intensidad de entrada ya que, por la ecuación de continuidad de la red anterior, la suma de intensidades conectadas al circuito precedente es cero. Del mismo modo, las variables de salida son la tensión y la intensidad de salida y sólo hay una intensidad de salida, por aplicación de la ecuación de continuidad a la red posterior. Esta separación exige la condición suplementaria de que **no haya ninguna conexión entre la red anterior y la posterior.**

De las cuatro variables que se observan desde fuera (figura 3-18) dos son interiores, e_1 y e_2, y dos son ligaduras exteriores, i_1 e i_2. La red aportará R + N - 1 ecuaciones y hay R + N - 1 variables interiores. Eliminando todas menos dos (e_1 y e_2) quedan **dos** ecuaciones con **cuatro** variables, e_1, e_2, i_1 e i_2. Como se trata de ecuaciones lineales y homogéneas, se pueden escribir tomando dos variables como dependientes y dejando las otras dos como independientes, escribiendo las ecuaciones en forma **matricial.**. Existen seis modos distintos de tomar dos variables entre las cuatro sin tener en cuenta el orden (combinación de cuatro elementos de dos en dos) , lo que da lugar a seis matrices de parámetros suministrados por la red.

Por ejemplo, si tomamos las dos tensiones como variables dependientes obtenemos la **matriz de impedancias**, correspondiente al sistema:

$$e_1 = z_{11}\ i_1 + z_{12}\ i_2$$
$$e_2 = z_{21}\ i_1 + z_{22}\ i_2$$

que en cálculo matricial puede expresarse como que el vector e_i es igual al producto de la matriz z_{ij} por el vector i_j .

Los parámetros coeficientes de la matriz de impedancias pueden definirse a la vista de las ecuaciones de la siguiente forma:

z_{11} es la impedancia de entrada con la salida abierta

(matemáticamente, de la ecuación, z_{11} es el cociente e_1/i_1 para $i_2 = 0$).

z_{12} es la impedancia de "transferencia" de entrada con la entrada abierta (e_1/i_2 para $i_1 = 0$).

z_{21} es la impedancia de transferencia de salida con la salida abierta (e_2/i_1 para $i_2 = 0$).

z_{22} es la impedancia de salida con la entrada abierta (e_2/i_2 para $i_1 = 0$).

Si resolvemos las ecuaciones anteriores (invertimos la matriz), se obtiene la **matriz de admitancias**, las ecuaciones toman la forma :

$$i_1 = y_{11}\ e_1 + y_{12}\ e_2$$
$$i_2 = y_{21}\ e_1 + y_{22}\ e_2$$

Los parámetros, que tienen dimensiones de admitancias, pueden definirse de modo análogo al anterior.

Si tomamos como variables dependientes las de salida, se obtiene la **matriz característica**, que es la de las ecuaciones:

$$e_2 = a_{11}\ e_1 + a_{12}\ i_1$$
$$i_2 = a_{21}\ e_1 + a_{22}\ i_1$$

Se trata de una matriz "híbrida" (los coeficientes no tienen todos las mismas dimensiones) que transforma el vector de entrada { e_1, i_1 } en vector de salida { e_2, i_2 }. Si se conectan dos o más cuadripolos sucesivamente se puede pasar de las variables de entrada del primero a las de salida del último **multiplicando** las matrices. Aunque este tratamiento generalizado es atractivo matemáticamente, se aplica en raras ocasiones.

La matriz inversa de la característica da el vector de entrada en función de el de salida y puede usarse de forma análoga a la directa.

Las dos matrices que faltan son las de mayor aplicación técnica. Escribamos en primer lugar la llamada de **parámetros híbridos** (las anteriores también eran híbridas, pero se reserva esta denominación para la que vamos a escribir ahora) :

$$e_1 = h_{11}\ i_1 + h_{12}\ e_2$$
$$i_2 = h_{21}\ i_1 + h_{22}\ e_2$$

El significado de los parámetros híbridos es:

h_{11} es la impedancia de entrada con la salida cortocircuitada.

h_{12} es la amplificación inversa con la entrada abierta.

h_{21} es la amplificación de corriente con la salida cortocircuitada.

h_{22} es la impedancia de salida con la entrada abierta.

Estos parámetros se aplican gran número de veces porque, como veremos, son adecuados para los transistores.

Por último, la matriz más aplicada es la de parámetros técnicos, que se obtiene tomando como variables dependientes e_2 e i_1 . Por su importancia le dedicaremos un apartado especial.

3.13.- Parámetros técnicos.

Observemos el cuadripolo de la figura 3-20, que suponemos cargado con una impedancia Z_L. El cuadripolo efectúa una **transformación** de la señal e_1 que **entra** en el cuadripolo (por ejemplo conectando un generador ideal e_1) a la señal que **sale** del cuadripolo hacia la carga. Escribiremos las ecuaciones del cuadripolo en la forma que llamaremos de **parámetros técnicos**, tomando como variables dependientes e_2 e i_1:

$$e_2 = Ae_1 - Z_s i_2$$
$$i_1 = \frac{1}{Z_e} e_1 + r i_2$$

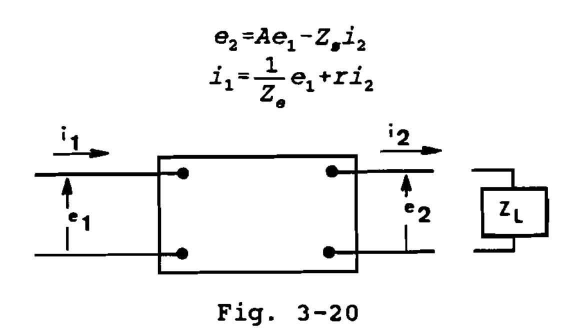


Fig. 3-20

A es la **amplificación**, a veces llamada **ganancia**, sin carga (el valor de e_2/e_1 para $i_2 = 0$).

Z_s es la **impedancia de salida** con la entrada cortocircuitada ($-e_2/i_2$ con $e_1 = 0$).

Z_e es la **impedancia de entrada** sin carga (e_1/i_1 con $i_2 = 0$).

r es el "efecto retroactivo" de i_2 sobre i_1 o amplificación de corriente con entrada cortocircuitada (i_1/i_2 para $e_1 = 0$).

Si la carga $Z_L \rightarrow \infty$ (sin carga, $i_2 = 0$) A es la amplificación y Z_e es la impedancia de entrada. En muchos circuitos se cumple esta aproximación, que es deseable en el sentido de que el funcionamiento del cuadripolo no depende de lo que se conecte a la salida.

La primera ecuación técnica es la ecuación de un generador de fuerza electromotriz Ae_1 e impedancia interna Z_s, o sea que el circuito del cuadripolo con entrada e_1 es **equivalente a un**

generador Ae_1 con impedancia Z_s. Por eso, a veces, a la primera ecuación técnica se le llama **de Thevenin**. La utilizaremos para estudiar la salida, como **ecuación de salida**. La figura 3-21 representa el circuito equivalente de Thevenin del cuadripolo con generador.

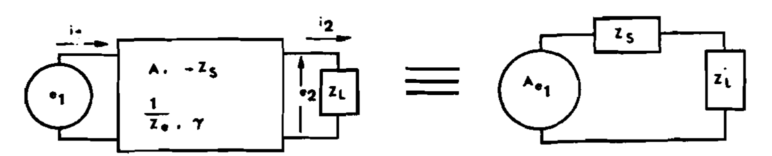


Fig. 3-21

La segunda ecuación expresa las condiciones **de entrada** en función de la carga. La utilizaremos para estudiar la carga que ve el generador en función de la carga.

Estudiemos ahora las condiciones del cuadripolo con **generador y carga**:

Si eliminamos i_2 entre la primera ecuación del cuadripolo y la ecuación de carga $e_2 = i_2\, Z_L$, o bien resolvemos directamente el divisor de tensión de la derecha de la figura 3-21, obtenemos:

$$\frac{e_2}{e_1} = A\frac{Z_L}{Z_L+Z_s}$$

que nos dice lo que vale la **amplificación con carga** y que es aproximadamente igual a la amplificación sin carga con el error relativo Z_s/Z_L.

La impedancia de entrada con la carga finita Z_L se obtiene eliminando i_2 y e_2 entre las ecuaciones del cuadripolo y la de carga. Sustituyendo $i_2 = e_2/Z_L$ en la segunda ecuación técnica se obtiene:

$$i_1 = \frac{1}{Z_e}e_1 + \frac{r}{Z_L}e_2$$

y sustituyendo ahora el valor de e_2 antes obtenido:

$$i_1 = \frac{1}{Z_e}e_1 + \frac{rA}{Z_L+Z_s}e_1$$
$$\frac{i_1}{e_1} = \frac{1}{Z_e} + \frac{rA}{Z_L+Z_s}$$

que nos dice que, con carga Z_L, aparece, en paralelo con la impedancia de entrada Z_e, otra que se llama **reflejada** y que vale:

$$Z_{refl.}=\frac{Z_L+Z_s}{rA}$$

Este valor será despreciable frente a Z_e si se cumple que puede despreciarse en la fórmula Z_s frente a Z_L y se cumple la condición $Z_{refl} \approx Z_L/rA >> Z_e$.

En la práctica es importante que la impedancia de entrada sea grande respecto a la del generador, para que éste, e_1, pueda considerarse como ideal y para esto son necesarias dos condiciones:

a) Que la impedancia de entrada Z_e sea grande respecto a la del generador.

b) Que la impedancia reflejada Z_L/rA también lo sea.

CAPITULO 4

CIRCUITOS PARTICULARES

4.1.- Introducción.

En este capítulo estudiaremos algunos circuitos pasivos que producen transformaciones útiles de las señales. En especial insistiremos en la resonancia y el **filtrado** basado en circuitos pasivos.

4.2.- Divisores de tensión.

La figura 4-1 representa un divisor de tensión compuesto por dos impedancias. Se trata de un cuadripolo elemental, cuyas variables de entrada son e_1, i_1 y las de salida e_2, i_2. Los terminales inferiores de entrada y salida son comunes y frecuentemente forman la tierra del sistema, aunque en principio no es necesario denominarla así.

Las ecuaciones que aporta la red son:

$$e_1 - e_2 = i_1 Z_1 \qquad \text{(rama } Z_1\text{)}$$
$$e_2 = (i_1 - i_2) Z_2$$

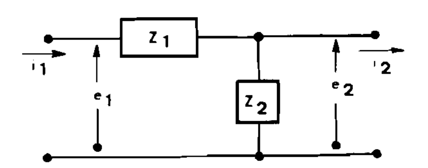


Fig. 4-1

En realidad ha habido una variable interna, la intensidad que circula por Z_2, que ha sido eliminada con ayuda de la ecuación del nudo Z_1 Z_2, al tomar como intensidad a través de Z_2 el valor $i_1 - i_2$.

Ordenando las ecuaciones se obtienen directamente los

parámetros híbridos:

$$e_1 = Z_1 i_1 + e_2$$
$$i_2 = i_1 - \frac{1}{Z_2} e_2$$

de donde $h_{11} = Z_1$, $h_{12} = 1$, $h_{21} = 1$ y $h_{22} = -1/Z_2$.

Eliminando i_1 entre las ecuaciones anteriores se obtiene la **primera ecuación técnica:**

$$e_2 = \frac{Z_2}{Z_1 + Z_2} e_1 - \frac{Z_1 Z_2}{Z_1 + Z_2} i_2$$

El divisor de tensión es equivalente a un generador de tensión:

$$\frac{Z_2}{Z_1 + Z_2} e_1$$

De aquí la denominación de "divisor", proporcional a la relación entre Z_2 y la suma $Z_1 + Z_2$. La impedancia interna es:

$$\frac{Z_1 Z_2}{Z_1 + Z_2}$$

que es el resultado de poner Z_1 y Z_2 en paralelo.

Este resultado podría haberse obtenido directamente aplicando el teorema de Thevenin al cuadripolo con un generador e_1.

Eliminando e_2 entre las ecuaciones del cuadripolo se obtiene la **segunda ecuación técnica:**

$$i_1 = \frac{1}{Z_1 + Z_2} e_1 + \frac{Z_2}{Z_1 + Z_2} i_2$$

La impedancia de entrada es $Z_1 + Z_2$. El parámetro r = A.

Si consideramos ahora el cuadripolo cargado con la impedancia Z_L, podemos aplicar todo lo que hemos deducido de

cuadripolos y , en particular :

La amplificación será $Z_2/(Z_1 + Z_2)$ si la impedancia de salida cumple $Z_1 Z_2 /(Z_1 + Z_2) << Z_L$.

El valor de la amplificación del divisor de tensión, dependiente de los valores de las impedancias, permite hacer que la amplificación sea una función de la frecuencia con determinadas propiedades, eligiendo convenientemente las impedancias. Basados en estos principios, veremos en los párrafos siguientes **filtros** con amplificación variable con la frecuencia en la forma deseada, obteniéndose esta dependencia con la elección adecuada de las frecuencias f_1 y f_2.

La impedancia de entrada $Z_1 + Z_2$ debe ser alta respecto a la impedancia o resistencia interna del generador e_1.

4.3.- Divisor R R.

El divisor de tensión formado por una resistencia y un "potenciómetro" (resistencia variable), representado en la figura 4-2, es un método sencillo y barato de obtener una fracción ajustable de una tensión determinada y encuentra muchas aplicaciones como control de volumen o para obtener una tensión variable.

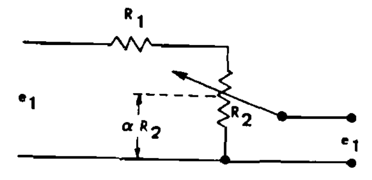


Fig. 4-2

Si llamamos αR_2 a la fracción de R_2 tomada por el potenciómetro, podemos decir:

a) La tensión de salida será:

$$e_2=\frac{\alpha R_2}{R_1+R_2}e_1$$

ajustable entre 0 y $R_2 e_1 /(R_1 + R_2)$.

b) La resistencia de entrada será $R_1 + R_2$.

c) La resistencia de salida será:

y su valor máximo ocurrirá cuando $\alpha R_2 = (1 - \alpha)R_2 + R_1 = (R_1 + R_2)/2$. y vale $(R_1 + R_2)/4$.

$$\frac{\alpha R_2[(1-\alpha)R_2+R_1]}{R_1+R_2}$$

Si $R_2<R_1$, el valor máximo es $R_1R_2/(R_1 + R_2)$.

En ambos casos, estos valores deben ser mucho menores que la impedancia de carga.

4.4.- Divisor RC.

Utilizando un condensador como impedancia Z_2 se obtiene una amplificación decreciente con la frecuencia, es decir un **filtro pasabajos**. La figura 4-3 representa el circuito o filtro RC. Recordando que la impedancia del condensador es $Z_2 = 1/j\omega C$ se obtienen los siguientes resultados:

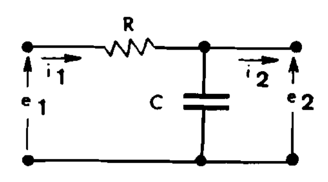


Fig. 4-3

a) La tensión de salida será:

$$e_2=\frac{\frac{1}{j\omega C}}{R+\frac{1}{j\omega C}}e_1=\frac{1}{1+j\omega RC}$$

su amplitud valdrá:

$$\frac{1}{\sqrt{1+\omega^2R^2C^2}}e_1$$

y su ángulo de fase respecto a e_1 será - arc tg ωRC.

b) La impedancia de entrada será :

$$R+\frac{1}{j\omega C},\, módulo \sqrt{R^2+\frac{1}{\omega^2C^2}}$$

y ángulo de fase ϕ = - arc tg $(\omega CR)^{-1}$

c) La impedancia de salida será:

$$\frac{\frac{R}{j\omega C}}{R+\frac{1}{j\omega C}}=\frac{R}{1+j\omega RC}$$

Las curvas correspondientes a e_2 / e_1 en función de la frecuencia en amplitud y fase están representadas en las figuras 4-4 y 4-5 y constituyen la llamada **respuesta de frecuencia en amplitud y fase**. Recuérdese que la frecuencia f es $f=\omega/(2\pi)$. La constante RC que caracteriza las curvas se llama **constante de tiempo** y es el inverso de la pulsación para la cual la **atenuación** del circuito (amplificación cuando es menor que 1) tiene un módulo $\sqrt{2}/2$ y una fase -45º. Esta pulsación y su frecuencia correspondiente, marcadas en la figura, se llaman **de corte**, y pueden también definirse en la curva como que la amplitud está a **- 3 dB del máximo**.

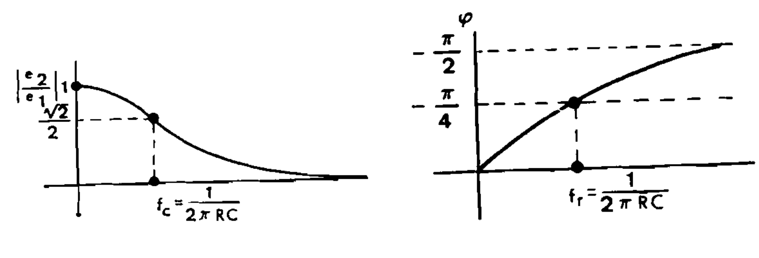


Fig. 4-4 Fig. 4-5

El circuito RC es un **filtro pasabajos**, las señales con frecuencia inferior a la de corte tienen atenuaciones inferiores a 3dB (la amplificación es superior a 0,7), y las frecuencias superiores a la de corte se atenúan más de 3dB (su amplificación es inferior a 0,7). En particular, la tensión continua (f→ 0) pasa sin atenuación, amplificación 1.

4.5.- Divisor CR.

El circuito dual del RC es el CR, formado como se ve en la figura 4-6 por un condensador y una resistencia. Aplicando las propiedades del divisor de tensión de forma análoga al párrafo anterior, se llega a los siguientes resultados:

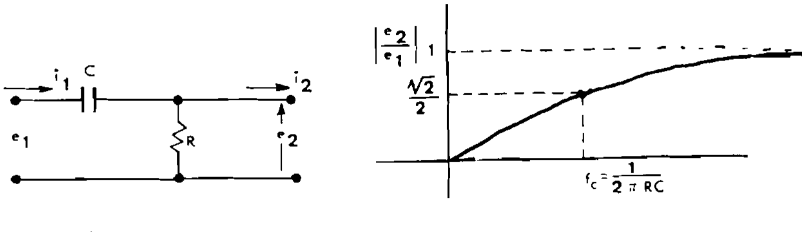


Fig. 4-6 Fig. 4-7

a) La tensión de salida sin carga será :

$$e_2 = \frac{R}{\frac{1}{j\omega C}+R} e_1 = \frac{j\omega CR}{1+j\omega CR} e_1$$

su amplitud es:

$$|e_2| = \frac{\omega CR}{\sqrt{1+\omega^2 R^2 C^2}} |e_1|$$

y el ángulo de fase:

$$\phi = \frac{\pi}{2} - arctg \frac{1}{\omega RC}$$

b) La impedancia de entrada sin carga es la misma del RC.

c) La impedancia de salida es la misma del RC.

La **respuesta de frecuencia** está representada en las figuras 4-7 y 4-8. La constante de tiempo es también RC y es el inverso de la pulsación de corte, a la cual la amplificación vale 0,7 y el desfase 45º.

El circuito CR es un **filtro pasaaltos**. Las frecuencias bajas están atenuadas con amplificación inferior a 0,7 y las altas pasan sin cambio, con amplificación superior a 0,7. En particular la tensión continua **no pasa**, debido al condensador.

4.6.- Divisor LRC.

Todavía más acusado es el efecto de filtro del divisor LC o LRC, siendo R la resistencia del bobinado de L (fig. 4-9).

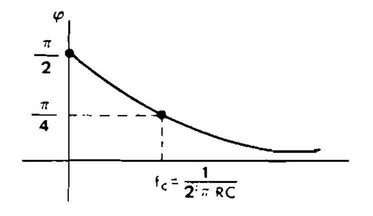


Fig. 4-8

L R e1 C e2

Fig. 4-9

a) La tensión de salida será :

$$\frac{\frac{1}{j\omega C}}{R+j\omega L+\frac{1}{j\omega C}}e_1=\frac{1}{1-\omega^2 LC+j\omega CR}e_1$$

b) La impedancia de entrada es :

$$R+j\omega L+\frac{1}{j\omega C}$$

c) La impedancia de salida es:

$$\frac{\frac{1}{j\omega C}(R+j\omega L)}{R+j\omega L+\frac{1}{j\omega C}}$$

Las curvas de **respuesta de frecuencia** están representadas en las figuras 4-10 y 4-11. y ponen de manifiesto el carácter de **filtro pasabajos**, aunque también puede usarse como pasabanda, dependiendo de lo agudo que sea la curva.

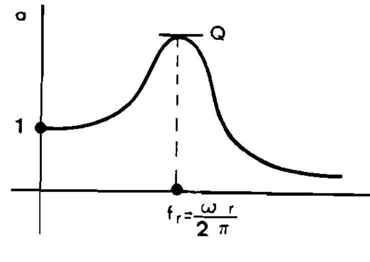


Fig. 4-10

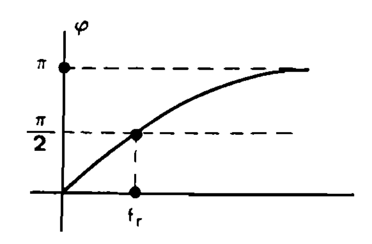


Fig. 4-11

La frecuencia de pulsación:

$$\omega_r=\frac{1}{\sqrt{LC}}$$

se llama **frecuencia de resonancia**. Suponiendo que R es pequeño frente a ω_r L, la amplificación tiene un máximo a ω_r de módulo:

$$Q=\frac{1}{\omega_r CR}=\frac{\omega_r L}{R}=\frac{1}{R}\sqrt{\frac{L}{C}}$$

Q se llama **factor de calidad** del circuito y representa lo

acusado de la resonancia (o lo pequeño que es R frente a la autoinducción).

4.7.- Resonancia LC.

El efecto de **resonancia** entre un condensador y una bobina se utiliza en muchas aplicaciones electrónicas en forma de **impedancia LC serie** (figura 4-12) y **impedancia LC paralelo** (figura 4-13). En ambas figuras se representa como resistencia exterior la de la bobina.

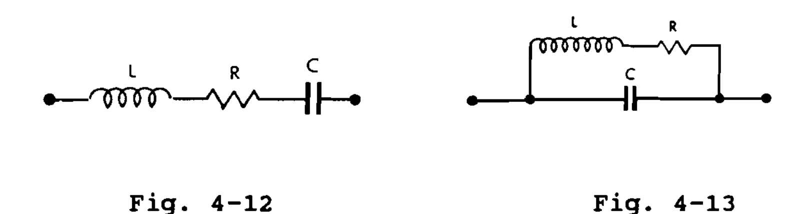


Fig. 4-12 Fig. 4-13

La impedancia de la resonancia serie LRC es :

$$Z_s=j\omega L+R+\frac{1}{j\omega C}$$

A la frecuencia de resonancia

$$f_r=\frac{1}{2\pi\sqrt{LC}}$$

las impedancias de la autoinducción y del condensador son iguales y de signo contrario y se anulan. El módulo de la impedancia es:

$$|Z_s|=\sqrt{R^2+[\omega L-\frac{1}{\omega C}]^2}$$

y pasa por un mínimo R en resonancia y es infinito para frecuencias 0 y ∞. En resonancia el valor de la tensión a través de la autoinducción es Q veces el de la aplicada, siendo

$$Q=\frac{\omega_r L}{R}=\frac{1}{R}\sqrt{\frac{L}{C}}$$

La fase de la impedancia es:

$$\operatorname{arctg}\frac{\omega L-\frac{1}{\omega C}}{R}$$

y pasa del valor $-\pi/2$ al valor $+\pi/2$.

El valor de la intensidad (módulo) para tensión aplicada unidad, en función de la pulsación es:

$$|i|=\frac{1}{\sqrt{R^2+[\omega L-\frac{1}{\omega C}]^2}}$$

La intensidad pasa por un máximo 1/R en resonancia y su representación es la de la figura 4-14.

Fig. 4-14

Para apreciar el ancho de la curva, calcularemos los valores de la pulsación en los puntos a - 3 dB del máximo, o sea a $\sqrt{2}/2$ del máximo, como se ve en la figura. Estos puntos son aquellos en que se cumple:

$$\omega_1 L-\frac{1}{\omega_1 C}=R$$

$$\omega_2 L-\frac{1}{\omega_2 C}=-R$$

Sumando esas dos expresiones se obtiene:

$$(\omega_1+\omega_2)L-\frac{\omega_1+\omega_2}{\omega_1\omega_2 C}=0$$

es decir $\omega_1\ \omega_2 = \omega_{r2}$, que nos dice que la pulsación de resonancia

es la media geométrica de las pulsaciones **de corte**. Para valores próximos, como son éstos, la media geométrica coincide aproximadamente con la aritmética.

Restando las expresiones anteriores se obtiene:

$$(\omega_1-\omega_2)L+\frac{\omega_1-\omega_2}{C\omega_1\omega_2}=2R$$

$$\omega_1-\omega_2=\frac{2R}{2L}=\frac{\omega_r}{Q}$$

es decir, la distancia entre las dos pulsaciones de corte es ω_r/Q. Esta distancia o ancho de la curva se llama **ancho de banda**. El ancho de banda es inversamente proporcional a Q.

La impedancia de la resonancia paralelo, de la figura 4-13 es:

$$Z_p=\frac{(R+j\omega L)\frac{1}{j\omega C}}{R+j\omega L+\frac{1}{j\omega C}}$$

Si suponemos de antemano que sólo nos interesan las frecuencias próximas a la resonancia y que el factor Q es elevado, podemos despreciar la R del numerador frente a $j\omega L$ y obtener:

$$Z_p=\frac{L/C}{Z_s}=\frac{Q^2R^2}{Z_s}$$

Como la figura 4-14 era la representación de $1/\,|Z_s|$, vale también para la impedancia paralelo **multiplicando la curva por** Q^2R^2. Así se observa:

El valor de la impedancia paralelo tiene un máximo en resonancia que vale **Q^2 R**. El ancho de banda vale ω_r / **Q**.

4.8.- Acoplamiento inductivo.

El acoplamiento inductivo o **transformador** está formado por dos bobinas colocadas de forma que cada una abrace un número de líneas de fuerza del campo magnético producido por la otra. El acoplamiento inductivo está representado en la figura 4-15 :

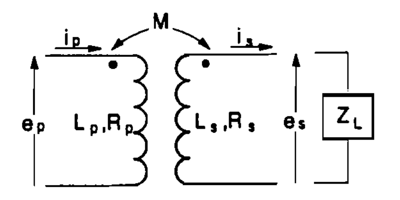


Fig. 4-15

Los puntos de la parte superior de la figura indican que entrando en el devanado de las bobinas desde ellos, el sentido de giro alrededor del núcleo es el mismo. El núcleo puede ser de aire, hierro magnético o ferrita (óxido de hierro aglomerado con una resina sintética). En estos dos últimos casos se dibuja una línea entre las bobinas. La bobina de entrada, representada a la izquierda, se llama primaria y la de salida, a la derecha, secundaria.

En la práctica se usan dos tipos distintos de acoplamientos:

En primer lugar, citemos el "transformador débilmente acoplado", generalmente con núcleo de aire o un pequeño núcleo de ferrita. Está caracterizado porque el secundario abraza sólo una pequeña parte de las líneas de fuerza del primario.

En segundo lugar, el "transformador fuertemente acoplado" se caracteriza porque prácticamente todas las líneas de fuerza producidas por el primario son abrazadas por el secundario. El núcleo es normalmente de un material de alta permeabilidad magnética, hierro "magnético" o ferrita, aunque excepcionalmente puede ser aire. Este transformador es adecuado para transformar potencia y es el usado en aplicaciones de Electrotecnia.

Suponemos que el transformador es conocido por el estudio de Electrotecnia. Recordemos que los parámetros que definen su comportamiento son :

L_p es la autoinducción de la bobina primaria, como si no existiese la secundaria. Por consiguiente, es la relación entre la tensión inducida en el primario y la derivada de la intensidad primaria, como cualquier autoinducción. La tensión inducida en el primario por la intensidad en el primario será $e_p = L_p \; d\, i_p/t$. En el plano $j\omega$, la tensión inducida en el propio primario por la corriente primaria será $j\omega i_p$.

R_p es la resistencia óhmica de la bobina primaria.

L_s es la autoinducción de la bobina secundaria.

R_s es la resistencia óhmica del secundario.

M es la **inducción mutua**. Es la relación entre la tensión inducida en el primario y la derivada de la intensidad secundaria. La tensión inducida en el primario por la intensidad secundaria será $e_p = M\, d\, i_s/t$. En el plano $j\omega$, la tensión inducida en el primario por el secundario es $j\omega\ M\ i_s$. Recordemos de Electrotecnia que la inducción mutua es también la relación entre la tensión inducida en el secundario y la derivada de la tensión primaria. En un transformador ideal es

$$M=\sqrt{L_p L_s}$$

A veces, en lugar de M se utiliza el coeficiente de acoplamiento ,

$$m=\frac{M}{\sqrt{L_p L_s}}$$

cuyo valor máximo es la unidad. En los transformadores "fuertemente acoplados" m=1.

Las ecuaciones del transformador de la figura 4-15, teniendo en cuenta que las intensidades simétricas que entran son i_p y $-i_s$ son :

Primario :

$$e_p-M\frac{d(-i_s)}{dt}=R_p i_p+L_p\frac{di_p}{dt}$$

Secundario:

$$e_s-M\frac{di_p}{dt}=R_s(-i_s)+L_s\frac{d(-i_s)}{dt}$$

Utilizando la transformación $j\omega$ serán:

$$e_p+j\omega M i_s=(R_p+j\omega L_p)\,i_p$$
$$e_s-j\omega M i_p=-(R_s+j\omega L_s)\,i_s$$

Estas son las ecuaciones del cuadripolo. Despejando i_p en la primera, se obtiene la segunda ecuación de parámetros técnicos:

$$i_p=\frac{e_p}{R_p+j\omega L_p}+\frac{j\omega M}{R_p+j\omega L_p}i_s$$

Sustituyendo en la segunda y despejando e_s se obtiene la primera ecuación técnica:

$$e_s=\frac{j\omega M}{R_p+j\omega L_p}e_p+[\frac{(j\omega M)^2}{R_p+j\omega L_p}-R_s-j\omega L_s]\,i_s$$

Los parámetros técnicos serán :

$$A=r=\frac{j\omega M}{R_p+j\omega L_p}$$
$$Z_E=R_p+j\omega L_p$$
$$Z_s=R_s+j\omega L_s+\frac{\omega^2M^2}{R_p+j\omega L_p}$$

Un caso particular muy importante es el del transformador **ideal**, que es un transformador fuertemente acoplado (m = 1) en el que además $R_p = R_s = 0$. En ese caso, los parámetros se convierten en :

$$A=\frac{M}{L_p}=\sqrt{\frac{L_s}{L_p}}$$
$$Z_E=j\omega L_p$$
$$Z_s=0$$

Normalmente se supone también que la autoinducción del primario es muy grande, y, con ella, la impedancia de entrada (infinito).

La impedancia reflejada con una carga Z_L será :

$$Z_R=\frac{Z_L}{Ar}=\frac{Z_L}{A^2}$$

El valor de A puede ponerse en función de los números de espiras del primario y secundario, recordando que las autoinducciones son proporcionales al cuadrado del número de espiras. Por consiguiente , si llamamos **n** a la relación entre el número de espiras del secundario y del primario, resultan las dos reglas importantes del transformador ideal :

$$A = n$$

$$Z_R = Z_L / n^2$$

Estas reglas caracterizan la aplicación del transformador ideal (o "fuertemente acoplado") como **adaptador de impedancias.** Cuando la carga es excesiva (Z_L demasiado bajo para ser mucho mayor que la impedancia de salida del circuito anterior) se coloca un transformador reductor intermedio (figura 4-16). La impedancia se refleja dividida por n^2, es decir, aumentada por ser $n<1$. El precio que se paga por ello es la pérdida de amplitud de la señal, que queda multiplicada por n.

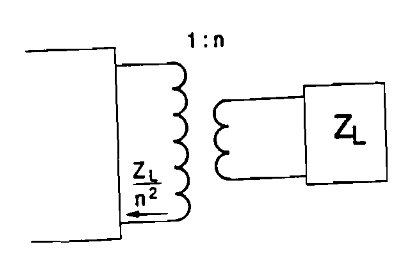


Fig. 4-16

4.9.- Acoplamiento doblemente sintonizado.

Un circuito muy utilizado por la forma de su respuesta de frecuencia de tipo **pasabanda**, que selecciona una banda o intervalo de frecuencias y atenúa las demás es el acoplamiento doblemente sintonizado o conjunto de dos circuitos resonantes acoplados entre sí, como se ve en la figura 4-17.

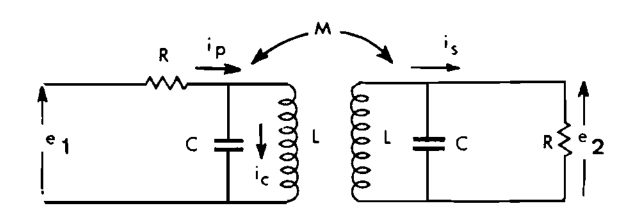


Fig. 4-17

Los dos circuitos resonantes son iguales, con bobinas de resistencia despreciable. Las resistencias exteriores R están conectadas como se ve en la figura. El coeficiente de acoplamiento será $m<<1$, supondremos que están débilmente acoplados. Las ecuaciones del transformador son ($e_s = e_2$):

$$e_p+j\omega Mi_s=j\omega Li_p$$
$$e_s-j\omega Mi_p=-j\omega Li_s$$

y las del resto del circuito:

$$i_c=j\omega Ce_p$$
$$i_s=j\omega Ce_s+\frac{e_s}{R}$$
$$e_1=e_p+R(i_p+i_c)$$

Eliminando i_c, i_p, i_s, e_p entre las cinco ecuaciones y

haciendo $R\sqrt{\frac{C}{L}}=q$; $\Omega=\omega\sqrt{LC}$ *se obtiene:*

$$\frac{e_1}{e_s}=\Omega^2[2m-\frac{2}{m}]+\frac{2}{m}+j\Omega^3[-\frac{q}{m}+qm]+j\Omega[\frac{1}{qm}+2\frac{q}{m}-\frac{m}{q}]+\frac{q}{j\Omega m}$$

es decir, la respuesta de frecuencia e_s / e_1 en amplitud es :

$$\frac{1}{\sqrt{[\Omega^2(2m-\frac{2}{m})+\frac{2}{m}]^2+[\Omega^3(qm-\frac{q}{m})+\Omega(\frac{1}{qm}+\frac{2q}{m}-\frac{m}{q})-\frac{q}{\Omega m}]^2}}$$

Esta fórmula, para valor constante de q, da lugar a una familia de curvas como la de la figura 4-18 :

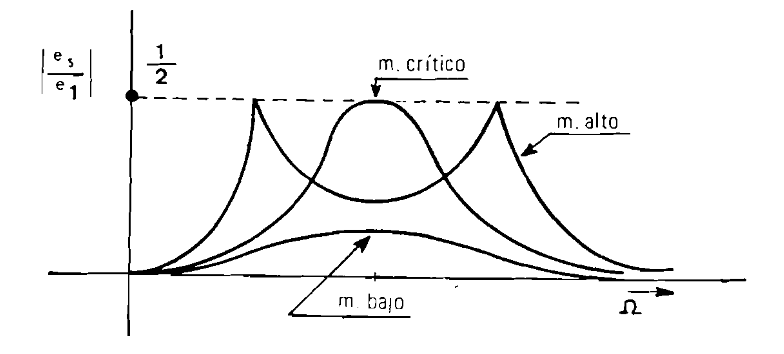


Fig. 4-18

La forma ideal de curva de filtro pasabanda ocurre para el valor **crítico** de m que hace que la curva no descienda en el centro. Para encontrar este valor, calculemos el valor de la respuesta de frecuencia en resonancia ($\Omega = 1$) y que es :

$$\frac{1}{\sqrt{4m^2+(qm+\frac{1}{qm}-\frac{m}{q})^2}}$$

Al variar q, el paréntesis del denominador, que se compone de dos sumandos cuyo producto es constante, se hace mínimo cuando son iguales, es decir:

$$qm=\frac{1}{qm}-\frac{m}{q} \text{ o sea } q^2m^2=1-m^2 \text{ o sea } m^2=\frac{1}{1+q^2} \text{ o sea } m\approx\frac{1}{q}$$

suponiendo q >> 1. Sustituyendo en la fórmula :

$$\left|\frac{e_2}{e_1}\right|=\frac{1}{\sqrt{4m^2+4q^2m^2}}\approx\frac{1}{2}$$

ya que m << 1 y qm vale aproximadamente 1.

En las proximidades de la resonancia, teniendo en cuenta que q = 1 / m >> 1 la fórmula queda :

$$\left|\frac{e_2}{e_1}\right|=\frac{1}{\sqrt{\frac{4}{m^2}(1-\Omega^2)^2+[\Omega^3+\Omega+\frac{1}{m^2}(-\Omega^3+2\Omega-\frac{1}{\Omega})]^2}}$$

o, con buena aproximación en las proximidades de $\Omega = 1$:

$$\left|\frac{e_2}{e_1}\right|=\frac{1}{\sqrt{4+\frac{16(1-\Omega)^4}{m^4}}}$$

Los puntos de amplitud - 3 dB es decir aquellos en que se verifica :

$$\left|\frac{e_2}{e_1}\right|=\frac{1}{2\sqrt{2}} \quad es\ decir\ 1-\Omega=\frac{m}{\sqrt{2}} \quad o\ sea\ \Omega=1\pm\frac{m}{\sqrt{2}}$$

Estos dos valores de Ω están a una distancia m √2, lo que significa que el ancho de banda es:

$$\Delta\omega=\frac{1}{\sqrt{LC}}m\sqrt{2}$$

siendo m el amortiguamiento **crítico,**

$$m=\frac{1}{q}=\frac{1}{R}\sqrt{\frac{L}{C}}$$

Recuérdese que q no es el valor de la calidad de resonancia, ya que R es una resistencia exterior que hace que q sea alto q>>1, aunque en la fórmula coincide con algo parecido a 1 / Q.

La figura 4-19 representa la respuesta de frecuencia para diferentes valores de q crítico (m = 1 / q).

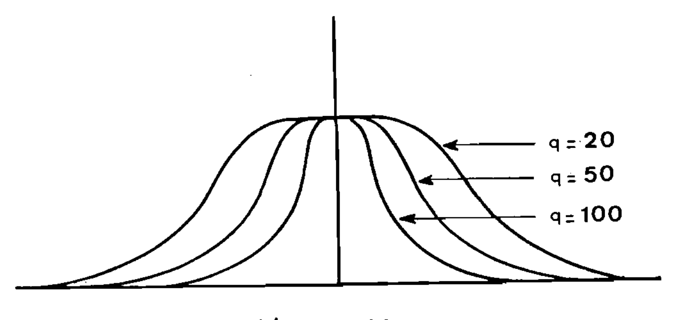


Fig. 4-19

Como puede apreciarse en la figura las frecuencias próximas al valor de resonancia pasan igualadas y las demás pasan fuertemente atenuadas. El **filtro pasabanda** es muy usado en receptores donde se querría un filtro de forma rectangular, aproximada por éste.

4.10.- Filtros pasivos.

En párrafos anteriores hemos visto diferentes circuitos que, debido a su respuesta de frecuencia, pueden usarse como **filtros** que modifican el espectro de la señal de entrada debido a la forma de su respuesta de frecuencia. En general el filtrado es una selección de respuesta de frecuencia para obtener los efectos

deseados. Podría decirse que la ecuación general del filtro es:

Espectro de entrada x Respuesta de frecuencia = Espectro de salida

Por ejemplo, una aplicación típica de los filtros es la eliminación de ruidos o señales no deseadas, basándose en que su frecuencia sea muy diferente a la frecuencia de la señal. Otra aplicación bien conocida es la atenuación de algunas frecuencias musicales para dar énfasis a otras.

Los filtros más sencillos son los basados en las propiedades del divisor de tensión, teniendo en cuenta la dependencia de la frecuencia de las dos impedancias que lo forman.

Los filtros más usados son los siguientes:

a) Filtros pasabajos.

Hemos visto ya el **filtro RC** (Fig. 4-3) y el **LC** (Fig. 4-9). La diferencia fundamental es que el descenso de la respuesta de frecuencia en el LC es más abrupto que en el RC, debido a la resonancia (lejos de la resonancia puede aproximarse con un descenso de segundo grado, en contraste con el primer grado del RC). Sin embargo la sencillez del RC hace que sea el filtro más usado en Electrónica.

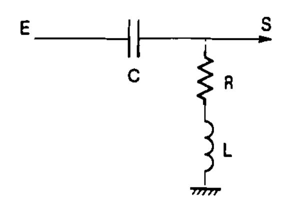
Podría pensarse en un filtro LR dual del RC, pero se descarta por ser más complicado y no tener ninguna ventaja.

b) Filtros pasaaltos.

Hemos visto el más sencillo y más usado, el **filtro CR** de la figura 4-6. Además de eliminar las frecuencias bajas, permite cambiar el nivel de continua por el efecto separador del condensador.

También aquí podemos decir que no se usa el dual, circuito RL, por no tener ninguna ventaja, y no separar la componente continua (sólo la eliminaría).

Sin embargo sí es útil el dual del LC, que sería el **CL**, representado en la figura 4-20, junto con su respuesta de frecuencia. La resistencia del bobinado de L se representa por separado. También aquí, la respuesta es más abrupta que la del CR.



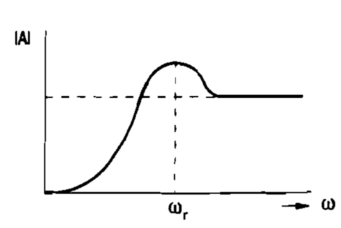


Fig. 4-20

c) Filtros pasabanda.

Hemos visto ya el acoplamiento inductivo doblemente sintonizado de la figura 4-17. Aunque es algo complicado, permite seleccionar una banda de frecuencias (correspondiente a su ancho) y presenta cortes abruptos de tipo resonante.

Utilizando las propiedades del divisor de tensión es fácil diseñar un filtro pasabanda colocando una resonancia paralelo en lugar de Z_2, como se ve en la figura 4-21.

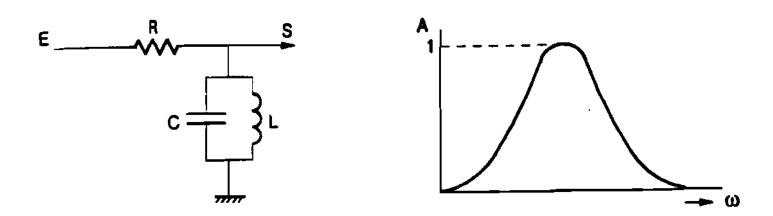


Fig. 4-21

En la misma figura se representa la respuesta de frecuencia que es similar a la variación de impedancia del circuito resonante, como era de esperar. Si se desprecia la resistencia de la bobina, la impedancia del circuito resonante a la pulsación o frecuencia de resonancia es infinito, con lo que la respuesta tiene ahí un máximo 1, descendiendo rápidamente a ambos lados.

El dual del anterior puede construirse con una impedancia serie en lugar de Z_1, presentando la misma respuesta de frecuencia.

Para reforzar más las propiedades del divisor de tensión, pueden usarse dos resonancias serie-paralelo, como se ve en el **filtro LC-CL** de la figura 4-22, donde también se representa la respuesta de frecuencia.

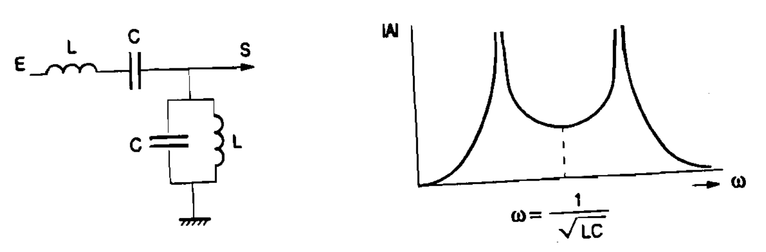


Fig. 4-22

Si se usan dos circuitos que tengan la misma frecuencia de resonancia, a ese valor el circuito serie tiene impedancia cero y el paralelo infinito, por lo que la fórmula del divisor de tensión da respuesta 1. A ambos lados de la resonancia común aparecen dos puntos singulares de respuesta infinito que a veces se llaman "de resonancia cruzada". Cuando los valores de las resistencias de las bobinas no son despreciables, la curva se

"suaviza", haciéndose más próxima a la del acoplamiento doblemente sintonizado.

d) Circuitos trampa

Son aquellos en que se desea suprimir una determinada frecuencia y dejar pasar las demás. Una vez más, el modo típico de conseguirlo es poner un circuito resonante en la impedancia adecuada del divisor de tensión.

La figura 4-23 es un **circuito trampa** con resonancia serie en Z_2. La respuesta de frecuencia está dibujada en la misma figura.

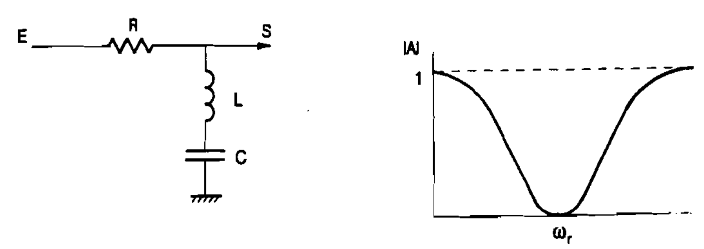


Fig. 4-23

Una típica aplicación del circuito trampa es la eliminación de perturbaciones procedentes de la red de suministro de energía, que son de una frecuencia bien determinada, por ejemplo 50 Hz. Otra aplicación típica es la eliminación de canales próximos a uno de comunicaciones que se quiera recibir, por ejemplo en televisión.

La técnica de elegir un filtro adecuado cuya respuesta de frecuencia se acerque lo más posible a la deseada se llama **Síntesis de filtros pasivos.** Existe una metodología para aproximar la respuesta de frecuencia a la deseada a base de polinomios que dan esa aproximaoión (polinomios de Tschebicheff) y que luego pueden realizarse con la elección adecuada de filtros sencillos y su combinación. Sin embargo, la enorme influencia de los elementos de filtro unos sobre otros (a través de sus impedancias finitas de entrada y salida) complica la composición de filtros a base de conectar filtros sencillos unos con otros. Con la aparición de los amplificadores operacionales, la adaptación se ha facilitado y la síntesis de filtros ha pasado a ser de **filtros activos, que estudiaremos más adelante**, quedando reducida la síntesis de filtros pasivos a la elección de alguno de los citados anteriormente.

CAPITULO 5

SEMICONDUCTORES

5.1.- Generalidades.

Los transistores fueron inventados por Bardeen y Brattain en 1.948. Pronto reemplazaron a los dispositivos de vacío, debido a sus ventajas fundamentales de solidez mecánica, tamaño reducido, mayor vida media y precio menor, a pesar de su peor comportamiento a altas frecuencias.

Las válvulas de vacío habían marcado el camino a seguir: encontrar un dispositivo que controlase una corriente eléctrica con muy poca o ninguna aportación de energía, Los dispositivos de material semiconductor, a lo que llamaremos también **semiconductores** consiguieron controlar la conductividad a través de la concentración de cargas "libres" en su seno. Ya se conocían dispositivos análogos de "cristal" que usaban el efecto de una punta metálica en contacto con un cristal de galena (sulfuro de plomo con impurezas de plata) y, sobre esa pista, se construyeron los primeros transistores. El control de cargas libres sustituyó al control por distribuciones de potencial, típico de las válvulas.

Existen actualmente muchos dispositivos "semiconductores" basados en las propiedades de los materiales del mismo nombre. En este capítulo trataremos de repasar algunos de los principios básicos en los que se basan y veremos su utilización en dispositivos concretos. Comenzaremos por las **bandas de energía**, esenciales para comprender el funcionamiento de los semiconductores, nivel de **Fermi** y su control, pasando luego a dispositivos concretos como diodos y transistores. Para el estudio de este capítulo son necesarios algunos conocimientos de **Física del Estado Sólido** y nociones de **Mecánica Cuántica**. Naturalmente haremos uso de ambas sin entrar en profundidad en ellas, omitiendo demostraciones o justificaciones que pueden encontrarse en la literatura especializada.

5.2.- Estructura atómica.

Utilizando la imagen clásica del átomo de Bohr, con núcleo

cargado positivamente y electrones en órbitas, podemos referirnos al "átomo de hidrógeno, con un núcleo y un electrón en órbita circular de radio r y calcular su energía total, suma de las energías potencial y cinética. Llamemos q a la carga del electrón (igual y contraria a la del núcleo), m a la masa del electrón y $4\pi\epsilon_0$ a la constante de Coulomb. La atracción f que ejerce el núcleo sobre el electrón será igual a la fuerza centrífuga,o sea:

$$f=\frac{q^2}{4\pi\epsilon_0 r^2}=\frac{mv^2}{r}$$

La energía cinética será:

$$\frac{1}{2}mv^2=\frac{q^2}{8\pi\epsilon_0 r} \text{ y la potencial } -\frac{q^2}{4\pi\epsilon_0 r}$$

La energía total es la suma de las dos y vale :

$$E=-\frac{q^2}{8\pi\epsilon_0 r}$$

igual en valor absoluto a la energía cinética. Tanto la energía total como la potencial son negativas y tienen un valor máximo (infinito) a distancia r=0. De esta fórmula se deduce que la energía del electrón, siempre negativa, depende solamente de la distancia al núcleo. Está pues justificado hablar de **niveles** de energía que corresponden a distancias. La dependencia de la energía es **hiperbólica** en función de la distancia, como se ve en la figura 5-1. Para alejar al electrón del núcleo hay que aportar energía, "subir" de nivel, hasta llegar al nivel cero (el máximo), cuando se aleja indefinidamente, es decir, se arranca del núcleo.

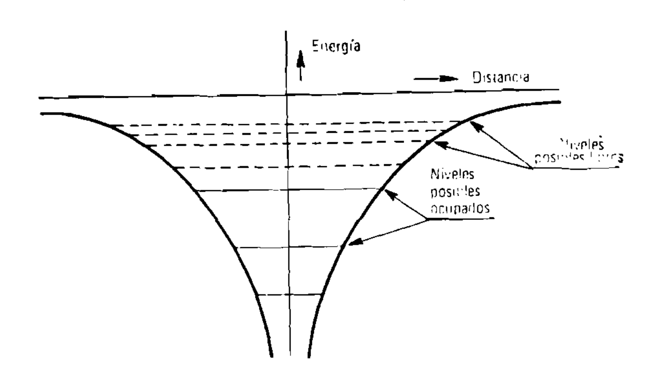


Fig. 5 -1

Según la teoría cuántica, no todos los niveles son **posibles**, sino sólo aquellos que corresponden a soluciones de la ecuación de Schrödinger. Entre esos niveles posibles **no** puede haber electrones. Por otra parte cada **nivel** (con una combinación de números cuánticos) sólo puede estar ocupado por un electrón (principio de exclusión de Pauli).

5.3.- Modelo con más de un átomo.

La presencia de un segundo átomo en las cercanías del primero influye poco en los electrones próximos al núcleo, ya que la energía suplementaria debida a la presencia del segundo átomo, a distancia grande respecto al volumen atómico, es despreciable. Cuando la distancia atómica es comparable con el radio de las órbitas (para los niveles superiores) hay un **desdoblamiento** de estos niveles y la energía puede tomar dos valores posibles ligeramente distintos. Los electrones del nivel exterior se "comparten" entre los átomos vecinos y su presencia en uno u otro átomo es aleatoria. Estos niveles compartidos forman una ligadura o **valencia** que mantiene los átomos unidos. Los niveles libres (por encima del último ocupado) también se desdoblan . A estos niveles se les llama, por razones que veremos después, niveles de **conducción**, mientras que a los niveles compartidos se les llama de **valencia**. La figura 5-2 muestra estos dos tipos de niveles.

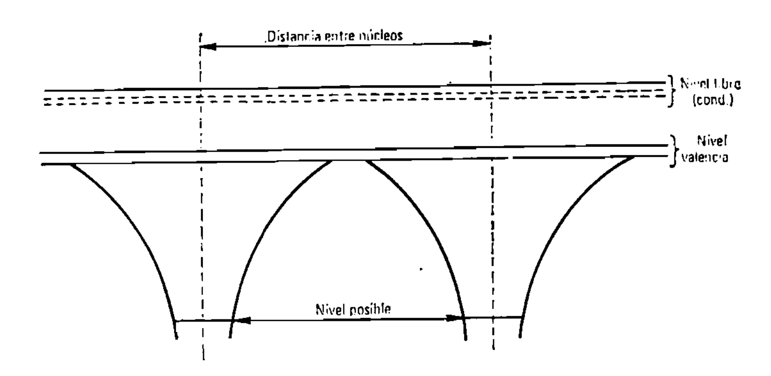


Fig. 5-2.

Cuando el número de átomos aumenta a distancias tales que los niveles de valencia se entrelazan, se forma un **cristal** y aparece una **banda de valencia** que contiene los electrones compartidos por átomos próximos y una **banda libre o banda de conducción**, vacía a bajas temperaturas, como se ve en la figura 5-3.

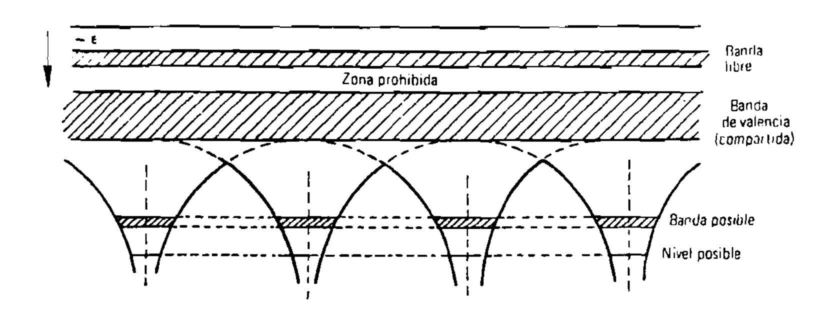


Fig. 5-3.

En esta figura, se dice que las abscisas representan "una dimensión en el seno del sólido", pero no tienen un significado matemático exacto, mientras que las ordenadas representan energía. El diagrama es unidimensional, si bien la dimensión horizontal da una indicación de distancias en la red cristalina.

A temperatura de cero grados absolutos, la banda de valencia está llena y la de conducción está vacía, en este modelo. A temperaturas superiores, en particular a temperatura ambiente, un número de electrones tiene energía suplementaria suficiente para saltar a la banda de conducción. Estas cargas son **libres** ya que tienen niveles que corresponden a una situación exterior al átomo. Estas cargas **libres** se mueven por el seno del sólido sin pertenecer a ningún átomo determinado.

5.4.- Conducción. Cargas libres.

El fenómeno de la conducción eléctrica ha recibido numerosas explicaciones que correspondían a "modelos" distintos del aquí descrito. Así la versión más extendida de la naturaleza de la corriente eléctrica que circula por un conductor metálico ha sido la de un flujo de electrones a través del sólido, retenido de alguna manera en su interior. Desde el principio se ha visto que no se podía explicar la enorme velocidad de la corriente (medida con transitorios), próxima a la de la luz y que más bien apuntaba a que la corriente era la "propagación" de una perturbación por cambio de potencial. Las perturbaciones viajan a la velocidad de las ondas (en este caso electromagnéticas), es decir, lo mismo que la luz.

La teoría moderna de la conducción es que ésta es debida a las cargas libres presentes, consiste en una propagación del exceso de electrones a través de un "gas" de electrones libres.

Tomemos como ejemplo un hilo conductor de metal, (cobre o plata) que según nuestro modelo está constituido por una red cristalina cargada positivamente (al perder los electrones libres) y una **nube** o gas de electrones retenida en el interior del sólido por la carga de la red, pero de movimientos libres, como las partículas de un gas, limitados sólo por los choques entre ellos. Si conectamos un extremo al polo negativo y otro al positivo de una batería, aparece un exceso de electrones

"presionados" por el potencial negativo que es el que se propaga hasta expulsar electrones por el lado positivo. Los electrones que entran no son los mismos que salen (como ocurría en la imagen de la conducción por flujo) aunque son indistinguibles, por uno de los principios de la mecánica cuántica. La propagación de ese "exceso de electrones" o ligero aumento de potencial es muy rápida, como cuando se empuja a un grupo de partículas por un lado y se mueven todas hacia el otro lado. Se mueven todas las partículas y se mide un flujo de movimiento en cualquier sitio, pero la distancia recorrida por cada partícula es muy pequeña.

La conducción puede también producirse por la presencia de **agujeros o huecos libres.** Estos agujeros se producen por la salida de cargas libres negativas de la banda de valencia. Cuando un electrón salta a la banda libre deja un "hueco" que es un átomo cargado positivamente. Este átomo atrae a los electrones de algún átomo próximo, que rellena el agujero, pero produce otro en ese átomo. Aparentemente se **desplaza** el agujero de un átomo a otro, y se comporta, a todos los efectos, como **carga libre positiva**. Estas vacantes libres en la zona de valencia sólo se ponen de manifiesto como tales vacantes o agujeros si están en número limitado, ya que si las órbitas externas están vacías no hay vacantes ni cargas libres positivas. La presencia de cargas libres positivas es típica de los **semiconductores**, como veremos en el próximo párrafo. La conducción por cargas libres positivas ocurre igual que la de los metales. El exceso de cargas positivas en el lado conectado al polo positivo (producido por absorción de electrones) se propaga a gran velocidad hasta el lado conectado al polo negativo, que se rellena otra vez con electrones del polo negativo. Ha aparecido, en el caso de los semiconductores un nuevo tipo de cargas libres, las **cargas libres positivas** y un nuevo tipo de intensidad eléctrica con cargas **portadoras positivas** y que va del más al menos, coincidiendo la corriente real con la convencional.

En los metales, la conducción es debida a los electrones o cargas **portadoras negativas**, que existen en gran abundancia. La resistividad es baja. La conductividad está limitada por los choques entre las partículas, y la resistividad **aumenta con la temperatura**, ya que al haber más densidad de carga hay más choques entre ellas. Los **metales** son **PTC**, su coeficiente de variación de la resistividad con la temperatura es positivo.

En los **semiconductores** la conducción eléctrica es debida a la presencia de **portadoras negativas** (electrones) y **portadoras positivas** (huecos) presentes en igual cantidad. Como la densidad de cargas es reducida, como veremos, la conducción está limitada por la densidad de cargas libres y aumenta con ésta. Por consiguiente los **semiconductores** son **NTC**, su resistividad disminuye con la temperatura y el coeficiente de variación de la resistividad con la temperatura es negativo.

Esta característica física, el signo del coeficiente de variación de la resistividad con la temperatura, permite distinguir, a una temperatura determinada, los conductores de los semiconductores. El coeficiente negativo de la variación de la

resistividad con la temperatura permite construir resistencias NTC con semiconductores que tienen diversas aplicaciones técnicas.

5.5.- Bandas de energía. Aislantes, semiconductores y conductores.

La dependencia de la conductividad de las cargas libres es la propiedad que va a prporcionar dispositivos de control y vamos a usar ampliamente en Electrónica. Por eso, nuestro interés en los sólidos va a centrarse en las **bandas de energía** de los cristales de determinados materiales, y, en particular, en la **zona prohibida** donde no hay cargas y que separa la **banda de valencia** de la **banda de conducción.** A cero grados Kelvin (-273ºC) la banda de valencia está llena y la de conducción vacía. A temperatura ambiente, un número determinado de electrones por unidad de volumen recibe energía necesaria para saltar de una banda a la otra y aparecen **cargas libres negativas** (electrones) en la banda de conducción y **cargas libres positivas** (huecos) en la banda de valencia, en cantidad igual. En adelante sólo representaremos esta zona y los bordes de las dos bandas, que es donde cuantificaremos el fenómeno.

La distancia que tienen que saltar los electrones para cambiar de banda es muy importante para dar una primera idea cuantitativa del fenómeno. En la figura 5-4 se representan tres tipos de sólidos según la magnitud de esa distancia:

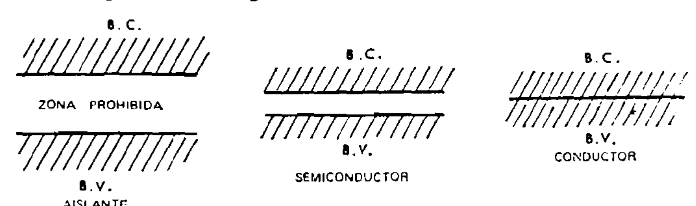


Fig. 5-4

En los **aislantes**, representados a la izquierda de la figura, la distancia (en energía) es tan grande que, con la agitación térmica, el número de electrones que la supera es despreciable. En estos sólidos no hay cargas libres ni movimiento de cargas y la **conductividad es cero.** Si se aplica un potencial exterior, los electrones "en exceso" suministrados por el polo negativo se quedan en los primeros átomos, y lo mismo ocurre con los electrones que salen al polo positivo. **La carga eléctrica es superficial** y no hay intensidad a través del sólido.Los materiales más usados como aislantes son los cerámicos y porcelanas.

En los **conductores**, representados a la derecha, la distancia entre bandas es cero o negativa (las bandas pueden superponerse) y el cristal está formado por una red atómica cargada positivamente y una **nube de electrones** libres. La banda de valencia (órbita externa) está vacía y no hay huecos ni cargas positivas móviles, **sólo** hay **cargas libres negativas**. La **conductividad** es muy **alta**, y está sólo limitada por los choques

entre las cargas, como ya hemos explicado. El cobre es el metal más usado como conductor, pero también lo son casi todos los metales, en particular la plata, el platino, etc...

En los **semiconductores**, representados en el centro, el salto de energía entre la banda de valencia y la de conducción es moderado, del orden de 1 eV, y el número de electrones que lo supera es muy dependiente de este salto, en cantidad relativamente pequeña (10^{-5}, por ejemplo), con lo que aparece un número determinado de **cargas libres negativas** y otro igual de **cargas libres positivas**, que determinan todas las propiedades de los semiconductores. La conductividad es moderada y depende de la concentración de cargas libres negativas o positivas. El material semiconductor más usado es el **silicio**, aunque también se usan otros como el germanio, selenio, etc... Estos materiales puros se llaman también **semiconductores intrínsecos**.

5.6.- Funciones de ocupación de los semiconductores.

Hasta ahora hemos hecho una descripción cualitativa de nuestro modelo de sólido. Para valorar cuantitativamente el fenómeno de aparición de cargas libres vamos a definir unas funciones de la energía que la caracterizan y que se llaman **funciones de ocupación**. En principio, nos referiremos a cristales semiconductores puros, normalmente de **silicio**, aunque también pueden ser de otros materiales como el germanio. Como ya hemos indicado, sólo nos interesan los valores de estas funciones a ambos lados de la zona prohibida en los semiconductores.

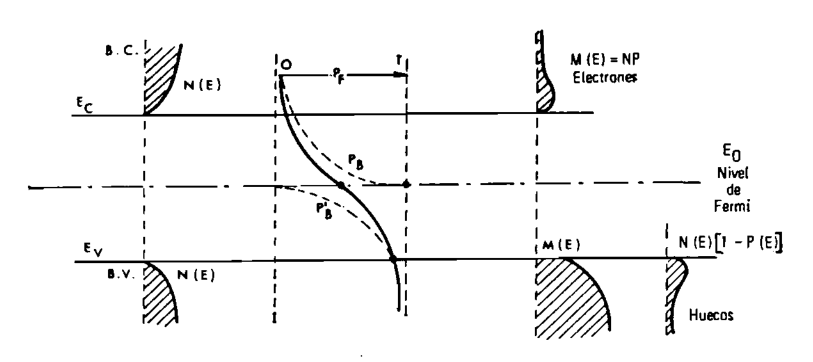


Fig. 5-5.

En la figura 5-5 están representadas estas funciones en un diagrama energético unidimensional, en el que las ordenadas representan siempre **energía** de las partículas en Ev y las abscisas representan el valor de las respectivas funciones.

La primera función, representada a la izquierda es la **densidad de niveles**, es decir, el número de niveles posibles por unidad de energía y unidad de volumen. Esta función describe cómo ha ocurrido el desdoblamiento de niveles al formarse el cristal y es una característica física propia del material. La llamaremos N(E). Recuérdese que es un número de electrones o huecos posibles por Ev y por centímetro cúbico. El número de electrones

"posibles" entre E y E+dE es N(E)dE cargas por c.c. Se trata pues de una densidad doble. Naturalmente, la función es simétrica (con algunas hipótesis simplistas) y crece hacia el interior de las bandas. El crecimiento resulta ser parabólico, proporcional a la raíz cuadrada de la distancia al borde. En la zona prohibida N=0 lo que corresponde a la definición de zona prohibida. Nosotros haremos la aproximación de tomar en lugar de N(E) un valor característico N en el interior (siempre próximo al borde) y 0 en la zona prohibida. Es decir, que sustituiremos N(E) por una forma rectangular. En realidad esta aproximación grosera no produce grandes errores debido a que el crecimiento es muy rápido en los bordes y aunque luego se separa bastante de la forma constante, siempre lo vamos a multiplicar por exponenciales que hacen que los valores alejados de los bordes no sean significativos.

La segunda función que vamos a estudiar es la **probabilidad de que un nivel esté ocupado**. El valor de la probabilidad, que es función de la energía del nivel y de la temperatura, varía entre cero (que significa nivel vacío) y 1 (que significa nivel lleno). La probabilidad de que un nivel esté ocupado viene dada por la **estadística de Fermi**, que es la siguiente:

$$P_F(E) = \frac{1}{1+e^{\frac{E-E_0}{KT}}}$$

en que T es la temperatura absoluta en grados Kelvin, K es la constante de Boltzmann K = 1,38 x 10^{-23} julios por grado y E_0 es la energía del **nivel de Fermi**, que es el nivel de probabilidad 0,5, ya que es fácil comprobar que P_F (E_0) = 0,5. El nivel de Fermi se encuentra con la condición de que la cantidad total de cargas positivas por unidad de volumen sea igual a la de cargas negativas. Posteriormente estudiaremos con más detalle la fijación del nivel de Fermi, aunque podemos adelantar que en los semiconductores intrínsecos está en el centro de la zona prohibida.

Como se ve fácilmente en la fórmula de Fermi, a temperatura tendiendo a cero, por encima del nivel de Fermi ($E-E_0>0$), la probabilidad tiende a cero, lo que corresponde perfectamente con la hipótesis de banda de conducción vacía. Por debajo del nivel de Fermi ($E-E_0<0$), la probabilidad tiende a 1, lo que corresponde a que la banda de valencia está llena.

Si consideramos la temperatura creciente a partir de cero, la curva de la estadística de Fermi se va separando de los ejes verticales 0 y 1 . En la banda de conducción, al separarse del eje 0 aparecen valores de probabilidad de que haya cargas negativas. En la banda de valencia, al separarse del valor 1 (vacantes ocupadas) va apareciendo una probabilidad 1 - P_F de que haya huecos, cargas positivas.

Si multiplicamos el número de vacantes por la probabilidad

de ocupación obtendremos la función M(E) = N(E) P(E) que es la esperanza matemática de ocupación y que supondremos que coincide con la ocupación real. La función M(E), función de **ocupación**, es la densidad de electrones por unidad de volumen y unidad de energía.

En el caso de la banda de conducción, la "ocupación" es la **densidad** (doble) de **cargas libres negativas** por unidad de energía y unidad de volumen. La densidad de cargas libres por unidad de volumen será la integral de la función M desde el borde de la banda hasta donde es despreciable y es el área rayada de la figura.

En el caso de la banda de valencia, la probabilidad de que una vacante esté libre es 1 - P_F y la esperanza matemática de la densidad de cargas positivas (huecos) es (1 - P_F)N(E). La ocupación de cargas positivas, dada por la fórmula anterior, es la **densidad** (doble) de cargas positivas por unidad de energía y unidad de volumen.

5.7.- Aproximación de Boltzmann.

La función:

$$P_F = \frac{1}{1+e^{\frac{E-E_0}{kT}}}$$

puede aproximarse, si $E-E_0$ es suficientemente grande, lo que prácticamente ocurre siempre en el interior de la banda, despreciando en el denominador 1 frente a la exponencial, con lo que resulta:

$$P_F \approx e^{-\frac{E-E_0}{KT}} = P_B$$

que es la "cola" o **estadística de Boltzmann** (también llamada de Maxwell- Boltzmann) para las cargas libres negativas, válida en la banda de conducción (por encima y relativamente lejos del nivel de Fermi).

La probabilidad de que haya un hueco en la banda de valencia es 1 - P_F , que puede escribirse:

$$1-P_F = \frac{e^{\frac{E-E_0}{KT}}}{1+e^{\frac{E-E_0}{KT}}} = \frac{1}{1+e^{-\frac{E-E_0}{KT}}}$$

Esta expresión tiene la misma forma que la probabilidad de

que haya un electrón, P_F, pero cambiando el signo a $E - E_0$. El valor de $E - E_0$, ahora negativo, sugiere un signo distinto para la energía para los huecos, tomando en ambos casos el **nivel de Fermi** como **origen o referencia de energía.**

Si, además, $-(E - E_0)$ positivo y grande, como ocurre en la banda de valencia, se puede despreciar de nuevo el 1 frente a la exponencial, quedando:

$$1-P_F=\frac{1}{1+e^{-\frac{E-E_0}{KT}}}\approx e^{\frac{E-E_0}{KT}}=P_B'$$

que es la **estadística de Boltzmann** o aproximación de la de Fermi para los huecos.

En la figura 5-5 se han representado con trazos las aproximaciones de Boltzmann para electrones y huecos, sobre la estadística de Fermi. Las curvas son una buena aproximación en las bandas y luego van a tomar valores 0 y 1 en el nivel de Fermi, lejos del valor correcto 0,5. En la zona prohibida las aproximaciones serían erróneas, pero como la densidad de niveles es cero, la probabilidad de ocupación no tiene ningún significado.

5.8.- Nivel de Fermi. Concentraciones de cargas libres.

Podemos calcular la cantidad total de cargas libres negativas por unidad de volumen con la integral:

$$n=\int_{E_C}^{\infty} N(E)\, P_F(E)\, dE$$

donde la integral se extiende entre el borde de la banda E_C e infinito, sustituyendo el borde superior de la banda por infinito, lo que facilita el cálculo de la integral y puede hacerse porque el integrando (P_F) es despreciable a partir de muy poca distancia del borde de la banda.

Si sustituimos N(E) por su valor aproximado N y P_F por la aproximación de Boltzmann, la integral se resuelve fácilmente:

$$n=\int_{E_C}^{\infty} N e^{-\frac{E-E_0}{KT}}\, dE=NKTe^{-\frac{E_C-E_0}{KT}}$$

Del mismo modo, podemos calcular la concentración total de cargas libres positivas p, obteniendo:

$$\int_{-\infty}^{E_V} N(E)\,[1-P_F(E)]\,dE=\int_{-\infty}^{E_V} Ne^{\frac{E-E_0}{KT}}\,dE=NKTe^{\frac{E_V-E_0}{KT}}$$

donde E_V es la energía en el borde de la banda de valencia y se ha supuesto el otro borde a distancia infinita.

La **ecuación que determina el nivel de Fermi** es **n = p**, o sea:

$$NKTe^{-\frac{E_C-E_0}{KT}}=NKTe^{\frac{E_V-E_0}{KT}}$$

La igualdad anterior se cumple si $E_C - E_0 = E_0 - E_V$, es decir, si el nivel de Fermi está en el centro de la zona prohibida.

La facilidad de las integrales anteriores sugiere también el cálculo de la concentración de cargas libres por unidad de volumen cuya **energía es superior al valor E_1**, dentro de la banda de conducción. Ese valor se llama $n(E_1)$ y es:

$$n_1=\int_{E_1}^{\infty} Ne^{-\frac{E-E_0}{KT}}\,dE=NKTe^{-\frac{E_1-E_0}{KT}}$$

Del mismo modo puede calcularse la concentración de cargas positivas cuya energía (en valor absoluto) es menor que E_1, y vale :

$$p_1=NKTe^{\frac{E_1-E_0}{KT}}$$

Por último, es interesante, como veremos después, calcular la **relación de concentraciones de cargas cuyos niveles son superiores a E_1 y E_2**, que vale :

$$\frac{n_1}{n_2}=\frac{NKTe^{-\frac{E_1-E_0}{KT}}}{NKTe^{-\frac{E_2-E_0}{KT}}}=e^{-\frac{E_1-E_2}{KT}}$$

Naturalmente, si $E_1 > E_2$ será $n_1 < n_2$, como se ve en la fórmula.

Por otra parte, la diferencia de energía total $E_1 - E_2$ puede expresarse utilizando la definición de potencial eléctrico ,

$$V=\frac{E_1-E_2}{q}$$

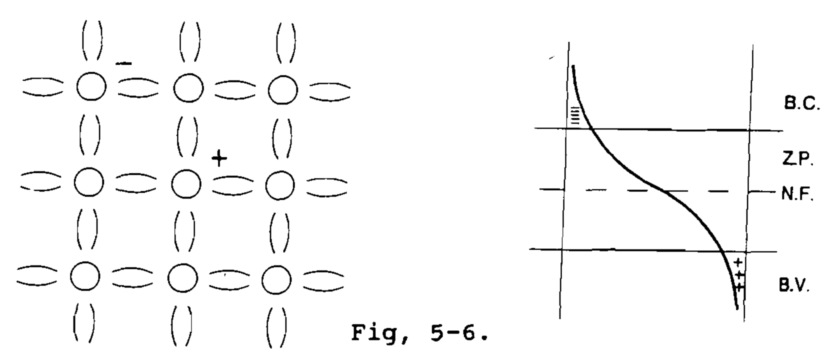
siendo q la carga del electrón. Sustituyendo, se obtiene:

$$\frac{n_1}{n_2}=e^{-\frac{qV}{KT}}=e^{-39V} \ ya \ que \ \frac{q}{KT}=39V^{-1} \ a \ 288^{\circ}K$$

Para las cargas positivas puede deducirse una fórmula análoga, cambiando el signo del exponente.

5.9.- Semiconductores intrínsecos y con impurezas.

El material semiconductor más usado es el silicio, que tiene cuatro valencias. En la figura 5-5 se representa una imagen plana del cristal en que se ve cómo cada átomo está ligado a los próximos por parejas de electrones compartidos. A temperatura ambiente alguno de estos electrones se ha liberado y se mueve en el sólido como carga libre negativa. El hueco que deja es también móvil y constituye una carga libre positiva. Constantemente se disocian las cargas dando una carga positiva y una negativa, y también hay un flujo de asociación en que desaparecen. Cuando un electrón cae a la zona de valencia (se asocia con un hueco) cede su energía a otro que salta a la banda de conducción. Las cargas son **indistinguibles** de forma que este flujo de asociación y disociación no es aparente, habiendo siempre la ocupación constante que hemos calculado.



Fig, 5-6.

A la derecha de la figura se ve la representación habitual del diagrama energético de un semiconductor **intrínseco**, como el silicio o el germanio. En él se aprecian las colas de Boltzmann, que son las colas de la estadística de Fermi. Estas mismas curvas, multiplicadas por N representan las densidades dobles de cargas por unidad de energía y unidad de volumen y multiplicadas por NKT representan las densidades de cargas libres cuya energía es superior a cada ordenada, para las cargas negativas o inferior para las positivas.

Veamos ahora lo que ocurre si el semiconductor se **dopa**. Se llama dopado a la introducción de una impureza en la red cristalina de un material de características muy parecidas y en muy pequeña proporción, para que la red no se altere.

Cuando este material tiene **un electrón más** en su órbita externa, como el arsénico, por ejemplo, se dice que la **impureza** es **donante**. El átomo donante ocupa el lugar de uno normal del semiconductor, forma ligaduras con los cuatro átomos próximos, usando cuatro de sus electrones de la órbita externa, y **cede** el quinto electrón como carga libre. Se produce así un **exceso** de cargas libres negativas (que, luego veremos, reduce el número de las positivas) y se queda cargado positivamente, porque le falta un electrón, creando así un nuevo tipo de carga positiva **fija**. (Fig. 5-7).

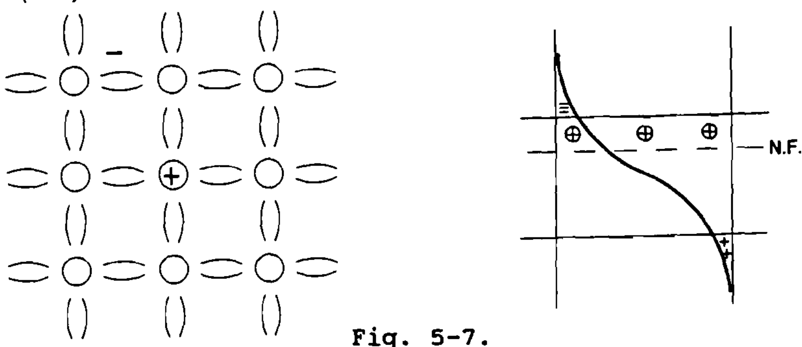


Fig. 5-7.

Este nuevo tipo de material, llamado **semiconductor tipo n** contiene así tres tipos de cargas eléctricas en su interior:

Cargas **libres negativas**, en gran cantidad, llamadas **mayoritarias**. Son electrones, como los ya conocidos en el vacío.

Cargas **libres positivas**, en pequeña cantidad, llamadas **minoritarias**. Son huecos y sólo existen en los semiconductores.

Cargas **fijas positivas**, que por primera vez aparecen como sujetas en el interior del sólido. Sólo existen en los semiconductores con impurezas.

Si llamamos n_l , p_l y p_f a las concentraciones por unidad de volumen de estas cargas, debe verificarse que:

$$p_f = n_l - p_l \text{ (carga total nula)}$$

Si tenemos en cuenta que, como veremos cuantitativamente, debido a la variación exponencial de las concentraciones, se verifica casi siempre que las cargas minoritarias son despreciables, la relación anterior queda $p_f = n_l$ y el modelo de semiconductor con impureza **donante** puede describirse como una distribución de cargas fijas equilibrando a una nube (ligera) de cargas libres. La Mecánica Cuántica se suele referir a este modelo como **caja de electrones libres**, ya que, ahora, se pueden retener los electrones en un recinto debido a las cargas fijas.

La figura 5-7 representa también el diagrama energético de un semiconductor de tipo n. En él se observan las cargas fijas, representadas con unos circulitos frente a las mayoritarias y las colas de Boltzmann, que en realidad mostrarían mucha mayor diferencia de la señalada en el diagrama.

Del mismo modo se puede usar una impureza **aceptante**, indio, por ejemplo, que tiene tres electrones en su órbita externa. La impureza sustituye a algunos átomos de la red cristalina y, para formar sus cuatro uniones roba un electrón de otro átomo generando así una carga libre positiva en exceso y quedando cargado negativamente.

En este tipo de semiconductores dopados de **tipo p** hay también tres tipos de cargas (figura 5-8):

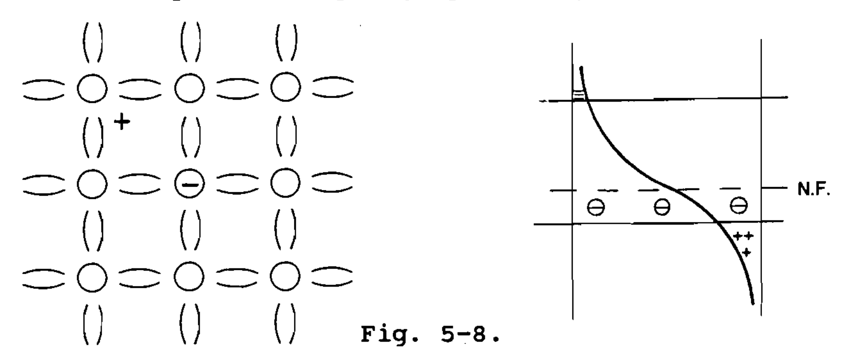


Fig. 5-8.

Cargas **libres positivas, mayoritarias**.

Cargas **libres negativas, minoritarias**, en muy pequeña proporción.

Cargas **fijas negativas**.

Se verificará que $n_f = p_l - n_l$. Aproximadamente será también $n_f = p_l$, despreciando las cargas minoritarias. Ha aparecido un nuevo tipo de cargas, electrones **fijos**, y el sólido es una **caja de huecos libres**, cargas positivas que se mueven en su interior.

Una aplicación inmediata de estas "cajas" de cargas libres es el llamado **efecto Hall**, que consiste en lo siguiente: Se origina un flujo de cargas portadoras (libres) en el seno de un semiconductor, o sea, una corriente eléctrica, mediante la aplicación de una diferencia de potencial exterior. Si se aplica un campo magnético perpendicular a la corriente, las cargas se desvían (ley de Lentz) y aparece una diferencia de potencial perpendicular al aplicado y al campo magnético, en el tercer eje. Esta diferencia de potencial permite medir el campo magnético.

5.10.- Nivel de Fermi en los semiconductores con impurezas.

Ya vimos que la igualdad de concentraciones de cargas positivas y negativas era la condición que fijaba el nivel de Fermi. En la figura 5-9 representamos las curvas de estadística

de Fermi multiplicadas por N, con lo que las curvas representan dentro de la banda de conducción densidades de cargas libres y el área respecto al eje 0 vertical representa la concentración de cargas libres negativas. En la banda de valencia, el área respecto al eje vertical N representa la concentración de cargas libres positivas. En la zona prohibida, la curva no tiene significado.

En la figura 5-9 puede verse cómo el nivel de Fermi se "adapta" automáticamente, bajando o subiendo, en los semiconductores de tipo p o n para que las áreas que deja la curva hasta los ejes verticales cumplan la condición de que su diferencia sea igual a la concentración de cargas fijas.

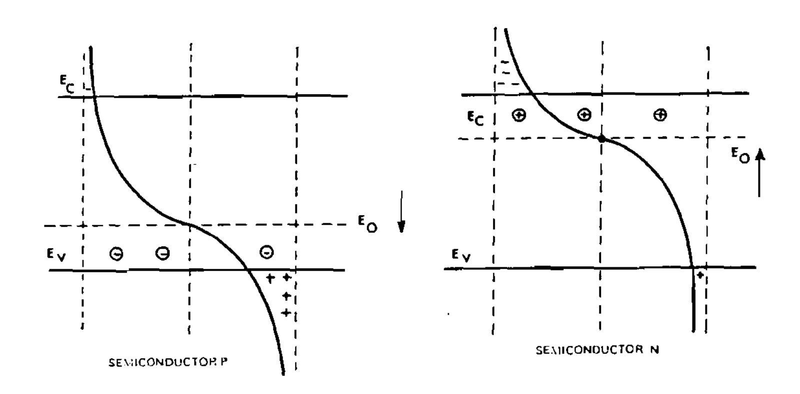


Fig. 5-9.

En la figura de la izquierda, se ve que el nivel **desciende** en los semiconductores de tipo **p**, y hay un gran número de cargas **mayoritarias positivas**, y un reducido número de **minoritarias negativas**. La diferencia de áreas es igual a la concentración de cargas **fijas negativas**.

En la figura de la derecha, se ve que el nivel **asciende** en los semiconductores de tipo **n**, y hay un gran número de cargas **mayoritarias negativas**, y un reducido número de **minoritarias positivas**. La diferencia de áreas es igual a la concentración de cargas **fijas positivas**.

En ambos casos se pueden aproximar las curvas en las bandas por la estadística de Boltzmann, pero entonces se pierde la imagen de la influencia del desplazamiento del nivel de Fermi sobre la diferencia de áreas.

Haciendo las hipótesis de simplificación que hemos hecho hasta ahora, la ecuación a cumplir por el nivel de Fermi será, para los semiconductores de tipo p:

$$NKTe^{-\frac{E_C-E_0}{KT}}-NKTe^{-\frac{E_0-E_V}{KT}}=n_f$$

Si hiciésemos $n_f = 0$, resultaría el nivel para semiconductores intrínsecos:

$$E_0=\frac{E_C+E_V}{2}$$

nivel que llamaremos E_i. Si llamamos ΔE a la **desviación** del nivel de Fermi, $E^i - E_0$, la ecuación que lo fija puede escribirse:

$$NKTe^{-\frac{E_C-E_i-\Delta E}{KT}}-NKTe^{-\frac{E_i+\Delta E-E_V}{KT}}=n_f$$

y, teniendo en cuenta que en un semiconductor intrínseco, la concentración de cargas libres (tanto negativas como positivas) $\mathbf{n_i}$, vale:

$$n_i=NKTe^{-\frac{E_C-E_i}{KT}}=NKTe^{-\frac{E_i-E_V}{KT}}$$

la ecuación queda:

$$n_ie^{\frac{\Delta E}{KT}}-n_ie^{-\frac{\Delta E}{KT}}=n_f$$

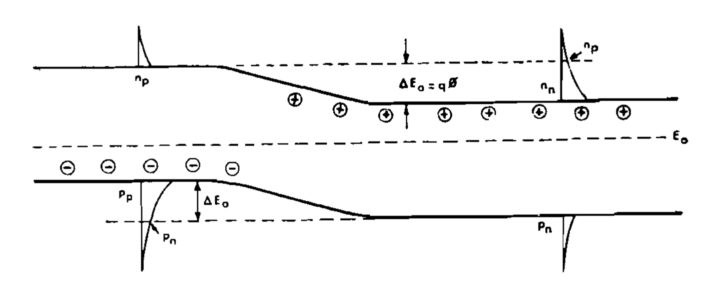
Esta ecuación permite calcular ΔE en función de n_f/n_i.

La ecuación es válida para semiconductores de tipo n, haciendo n_f negativo, con lo que resulta ΔE también negativo, como era de esperar.

5.11.- Contacto entre dos sólidos. La unión p-n.

Cuando dos sólidos se ponen en contacto, hay un intercambio energético hasta que se igualan la **temperatura y** el **nivel de Fermi**. En cuanto al nivel de Fermi, podemos justificar este axioma suponiendo que, al igualarse el nivel de Fermi, son iguales las concentraciones de cargas libres del mismo nivel de energía, ya que estas concentraciones dependen sólo de la distancia al nivel de Fermi, y, al ser iguales las concentraciones no hay difusión de cargas. Suponemos que no hay ninguna fuente exterior de energía que perturbe el equilibrio.

La figura 5-10 representa el diagrama energético de la **unión p-n** sin potencial exterior aplicado. Suponemos que la transición del material p al n es rápida e ideal, es decir, gradual y sin perturbaciones de la red cristalina, y también sin ningún material extraño entre ambos lados. Veamos los detalles de este diagrama energético:



a) UNION p-n SIN POTENCIAL EXTERIOR APLICADO

Fig. 5-10.

La parte de la izquierda es un semiconductor tipo **p**. Tiene, por consiguiente, cargas libres positivas mayoritarias y cargas libres negativas minoritarias, representadas por las "colas" de Boltzmann, cuya diferencia está equilibrada eléctricamente por las cargas fijas negativas, representadas frente a la banda de valencia. El nivel de Fermi está descendido, o sea, el resto elevado, si tomando el nivel de Fermi como referencia. Es conveniente recordar el significado las colas de Boltzmann, que tiene las siguientes tres interpretaciones:

a) La probabilidad de ocupación en función de la energía. Su valor es :

$$P_B = e^{-\frac{E-E_0}{KT}}$$

para las cargas negativas, tramo superior, y :

$$P_B' = e^{-\frac{E_0-E}{KT}}$$

para los huecos.

b) La densidad de niveles por unidad de volumen y unidad de energía, de valor NP_B o NP'_B. Es la misma curva, con factor N.

c) La concentración de cargas por unidad de volumen cuya energía (distancia al nivel de Fermi) es mayor que la del nivel E, de valor:

$$n(E)=NKTP_B \text{ o bien } p(E)=NKTP'_B$$

Las concentraciones **totales** por unidad de volumen son los valores anteriores particularizados para los **bordes** de las bandas, es decir:

$$n_p=NKTe^{-\frac{E_C-E_0}{KT}} \quad y \quad p_p=NKTe^{-\frac{E_0-E_V}{KT}}$$

Además se verifica que $p_p >> n_p$ y $p_p - n_p = n_f$. Las concentraciones totales son pues los valores de las bases de las curvas. Hay que hacer notar que el valor de la curva de cargas libres positivas a la altura de la banda de valencia del lado n es igual a ambos lados, es decir, p_n.

Todos estos valores a), b) y c) dependen exclusivamente de la distancia al nivel de Fermi (y de la temperatura) y **son iguales a ambos lados de la unión**, para cada valor de la energía, excepto cuando este valor cae en zona prohibida a uno de los lados, en cuyo caso es cero a ese lado.

La parte derecha de la figura es un material semiconductor de tipo **n.** El nivel de Fermi está elevado, lo que significa que el resto está descendido respecto a él. En esta parte existirán cargas libres mayoritarias negativas, cargas libres minoritarias positivas y cargas fijas positivas. Los valores del borde representan n_n y p_n. Hacemos notar que el valor de la curva de cargas libres negativas de la zona n a la altura de n_p es **también n_p.**

Utilizando ahora la última fórmula del párrafo 5.8, que da la relación de cargas cuyas energías son superiores a dos niveles de diferencia de potencial V, se obtiene :

$$\frac{n_p}{n_n}=\frac{p_n}{p_p}=e^{-\frac{q\phi}{KT}}=e^{-39\phi}$$

siendo ϕ la "barrera" o diferencia de potencial a ambos lados del borde de la banda de conducción, que es la suma de las desviaciones del nivel de Fermi debidas a las impurezas.

Por último, podemos subrayar que en la figura que existe un equilibrio eléctrico tanto en sentido vertical (a cada lado, como ya hemos visto), como horizontal. Puede interpretarse que la "batería" formada por las cargas libres mayoritarias (o por los excesos de las mismas) y que son móviles y tenderían a recombinarse, está equilibrada por la "batería" igual y contraria formada por las cargas fijas, que, en cierto modo, frenan ese movimiento con una especie de **barrera de potencial** formada por

estas cargas fijas.

Sin embargo, hay que prevenir contra la explicación demasiado elemental de esa simple barrera. Sin la zona prohibida y la identificación de la "barrera" con el salto debido a los desplazamientos del nivel de Fermi, no es posible explicar el efecto del diodo.

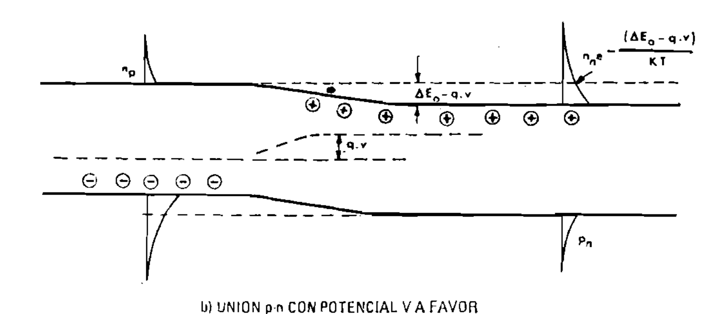
5.12.- Ecuación del diodo.

Un diodo es un componente pasivo formado por una unión p-n, es decir, una transición rápida entre un semiconductor dopado de tipo p y otro de tipo n.

Vamos a calcular la corriente eléctrica que fluye a través de la unión cuando se aplica una diferencia de potencial exterior V. Esta diferencia de potencial exterior se llama también **polarización**, y se dice que es **directa** o **a favor** cuando se aplica un potencial positivo al lado p y negativo al n o **inversa** o **en contra**, cuando se aplica un potencial positivo al lado n y negativo al lado p. Llamaremos positivo a V cuando esté a favor de la unión y negativo cuando esté en contra.

Utilizaremos simplificaciones de cálculo que no afectan al fenómeno físico (según Hunter), especialmente explicando el fenómeno con la difusión de cargas y suponiendo que esta difusión es proporcional a la diferencia de concentraciones. Un cálculo riguroso puede consultarse en el trabajo de Shockley, Sparks y Teal, Phis. Revue, julio 1.951.

Hemos visto ya la situación de equilibrio cuando no hay polarización (fig. 5-10). La distribución de cargas es igual a ambos lados de la unión por encima de la banda de conducción de la zona p. En el lado n hay una gran cantidad de cargas libres negativas que no tienen cargas libres al mismo nivel al otro lado, pero no pueden difundirse en la zona p porque estos niveles están en zona prohibida.



b) UNION p-n CON POTENCIAL V A FAVOR

Fig. 5-11.

Supongamos ahora que aplicamos una diferencia de potencial **a favor** o sea una **polarización directa** al diodo de la figura 5-10. El potencial exterior aplicado varía las referencias a

ambos lados de la unión, elevando el nivel de Fermi y , con él, todo el diagrama energético, en el lado n y bajando el nivel de Fermi y todo el diagrama en el lado p. La figura 5-11 representa el diagrama energético del diodo con polarización directa.

El salto del nivel de Fermi es ahora qV, siendo V la diferencia de potencial o tensión exterior aplicada. Enseguida se aprecia en la figura un gran exceso de cargas libres negativas mayoritarias del lado n sobre la pequeña concentración de cargas minoritarias del lado p, y **a niveles posibles**. Es evidente que puede haber una **difusión de cargas mayoritarias en la zona minoritaria**, produciéndose una corriente eléctrica debida a este flujo de cargas. Las cargas negativas mayoritarias que se difunden en la zona minoritaria son reemplazadas por electrones procedentes del polo negativo y las cargas que entran en la zona minoritaria son absorbidas por el polo positivo. Del mismo modo, las cargas mayoritarias positivas se difunden también en la zona minoritaria del lado n, contribuyendo de la misma forma a la corriente eléctrica.

El escalón del borde de la banda de conducción entre el lado n y el p está disminuido y es :

$$\Delta E_0 - qV = q\ (\ \phi - V\ )$$

La concentración de cargas libres negativas cuyo nivel es superior al borde de la zona positiva, puede calcularse por su relación con n_n y n_p y vale:

$$n_n e^{-\frac{q\phi - qV}{KT}} = n_p e^{\frac{q\phi}{KT}} e^{-\frac{q\phi - qV}{KT}} = n_p e^{\frac{qV}{KT}}$$

El exceso de cargas libres mayoritarias por encima de la zona minoritaria sobre n_p es :

$$n_p e \frac{qV}{KT} - n_p = n_p (e^{\frac{qV}{KT}} - 1)$$

Análogamente, para las cargas positivas o huecos hay un exceso de cargas mayoritarias al nivel de la zona minoritaria que vale:

$$p_n (e^{\frac{qV}{KT}} - 1)$$

Las corrientes de difusión son proporcionales a estos excesos y las constantes de proporcionalidad valen (ver Ebers y Moll, Proc. IRE, dic. 1.954) :

$$\frac{qD_n}{L_n} \quad y \quad \frac{qD_p}{L_p}$$

La corriente total será :

$$I=\left(\frac{qD_n}{L_n}n_p+\frac{qD_p}{L_p}p_n\right)\left(e^{\frac{qV}{KT}}-1\right)=I_0\left(e^{\frac{qV}{KT}}-1\right)$$

Esta fórmula puede aplicarse exactamente igual para V<0, es decir, con polarización **inversa**. El diagrama energético es el de la figura 5-12. La fórmula se convierte en $I = - I_0$, constante e independiente de V, que corresponde, como se ve en la figura a la difusión de **cargas minoritarias en la zona mayoritaria**. El valor de I_0 es muy pequeño, casi siempre despreciable. El dispositivo **conduce** cuando la tensión aplicada es **a favor** de la unión (polarización directa) y **no conduce** (excepto por la despreciable corriente de minoritarias que se difunden en zona mayoritaria) cuando la tensión aplicada es **en contra** (polarización inversa) de la unión.

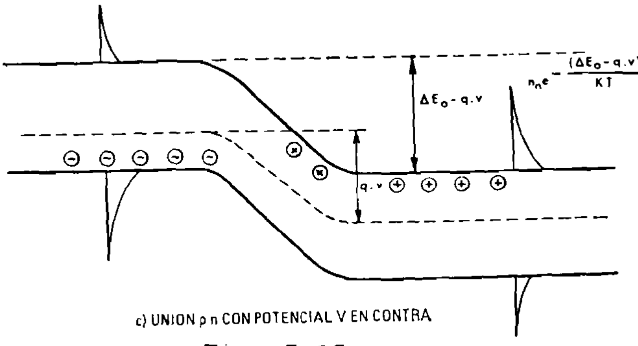


Fig. 5-12

5.13.- Característica del diodo.

El diodo es un componente de dos terminales o **electrodos** que corresponden a la conexión exterior al lado p de la unión, llamada **ánodo** y la conexión externa al lado n llamada **cátodo**. Esta terminología procede de las antiguas **válvulas**, que trataremos en un apéndice.

El símbolo del diodo es el representado en la figura 5-13, en que el ánodo está representado con una flecha que indica la entrada de la corriente convencional y el cátodo por una trazo perpendicular a la conexión. La **ecuación del diodo** es la que hemos deducido:

$$I=I_0(e^{39V}-1)$$

donde se ve que la corriente es una función que crece exponencialmente con V en sentido directo y es prácticamente constante para tensiones negativas (la exponencial es despreciable para valores negativos de V, para V=-1 sería e^{-39}) e igual a la **corriente inversa** I_0 . El coeficiente 39 V^{-1} es el valor de q/KT a 25ºC .

La representación gráfica de esta ecuación se llama **característica** del diodo y está representada en la figura 5-13. Como se ve en la figura, y también en la fórmula, I=0 para V=0, y la curva tiende asintóticamente a $-I_0$ para V negativo. Este valor es muy pequeño, del orden de 10^{-8} A. para los diodos de silicio. Este valor es creciente con la temperatura (más de lo que se aprecia en nuestras fórmulas debido a las simplificaciones realizadas), por lo que también lo es I, a pesar del decrecimiento de q/KT.

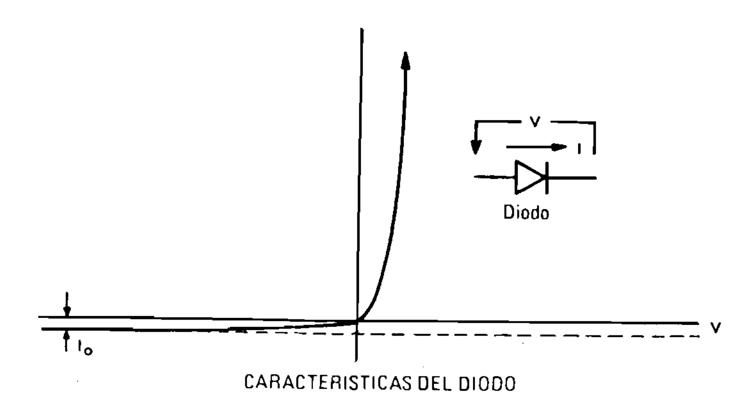


Fig. 5-13.

El crecimiento en la zona positiva es rapidísimo a partir de unas décimas de voltio. Por ejemplo, para I_0 = 1 nA., la intensidad I = 1 A. para V= 0,53 voltios. La curva puede aproximarse (despreciando el 1 en la fórmula) por :

$$I=I_0e^{39V}$$

que se multiplica por e cada vez que 39V aumenta 1, es decir, aproximadamente, **la intensidad se triplica cada 25 milivoltios.**

Estos dos fenómenos hacen que pueda a menudo simplificarse el comportamiento (**no lineal !**) del diodo por el sentido de "válvula" o "conmutador" que **en una dirección conduce** (con un potencial positivo infinitesimal aplicado) **y en la otra no** (aunque se aplique cualquier potencial). El gráfico se sustituye así por dos semirectas pegadas a los ejes (fig. 5-14), o a la

vista de la característica, por una semirecta horizontal y una vertical que deja una **tensión residual** de unas décimas de voltio, o una semirecta horizontal y otra inclinada que deja una **resistencia equivalente** en la zona de conducción.

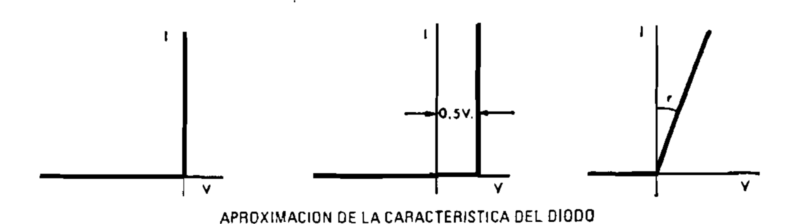


Fig. 5- 14

5.14.- Recta de carga.

La característica del diodo es una relación no lineal entre V e I. La característica deja un grado de libertad que debe determinarse por la red exterior, que debe aportar la ecuación que falta.

Un caso típico es el de un circuito con un diodo y una resistencia R al que se aplica una tensión exterior E.

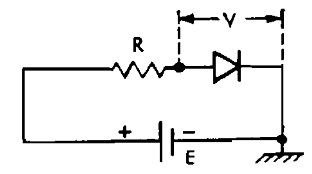


Fig. 5-15.

El circuito está representado en la figura 5-15. Las ecuaciones del circuito son la característica del diodo, dada aquí en forma gráfica en la figura 5-16 y la ecuación de la resistencia, que es:

$$I=\frac{E-V}{R}$$

La ecuación de la resistencia es la de una recta en el diagrama V-I de la figura 5-16. Esta recta llamada **recta de carga** corta a los ejes en los puntos V = E e I = E/R. La solución de valores V,I es la de los valores del punto de intersección de ambas curvas.

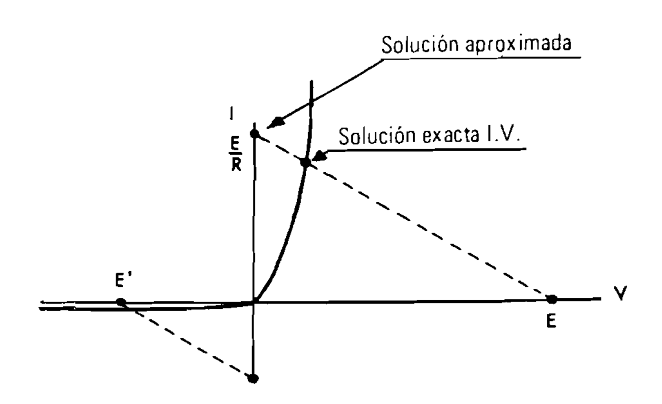


Fig. 5-16.

Si hubiésemos supuesto que el diodo era ideal, la solución sería 0, E/R, como se ve en la figura o se deduce directamente del circuito si se supone que el diodo "conduce", está cortocircuitado.

Si E es negativo, E=E', la intersección es prácticamente $I=-I_0$, V = E', como también se ve directamente considerando que el diodo no conduce, es un circuito abierto.

La recta de carga se mueve paralelamente a sí misma al variar el parámetro E y cambia de pendiente al variar R.

5.15.- Resistencia diferencial.

Al valor dV / dI se la llama **resistencia diferencial** y es la resistencia que presenta el diodo a una pequeña señal **incremental**, es decir, superpuesta a los valores de un punto de la característica, que se llama **punto de trabajo**. De la ecuación del diodo deducimos:

$$\frac{dI}{dV}=39\,I_0 e^{39V}=39\,(I+I_0)\approx 39\,I \quad o\ sea \quad \frac{dV}{dI}=\frac{1}{39\,I}$$

Esta variación de la resistencia con la intensidad (o también con la tensión, función de ella) significa que la resistencia para pequeñas señales superpuestas puede controlarse con la **polarización** o punto de trabajo. El valor cerca del origen es muy grande, de varios megohmios, pasando a ser de sólo ohmios para valores grandes de la intensidad. En esos puntos, hay que sumar la resistencia óhmica de las zonas p y n, no consideradas en el cálculo.

5.16.- El transistor npn. Ecuaciones.

El transistor **bipolar** es un componente **activo** (como veremos

hay que polarizarlo desde el exterior) compuesto por **dos uniones** (de ahí el nombre de bipolar o "bijunction") muy próximas entre sí, que, como veremos, influyen una sobre la otra e incluso controlan una a la otra. La figura 5-17 representa el diagrama energético de un transistor npn, formado por una zona de tipo n, otra intermedia de tipo p y otra de tipo n, con la zona intermedia muy estrecha. En estos diagramas, las ordenadas representan niveles energéticos y las abscisas "una dimensión en el seno del sólido", es decir, que, geométricamente, las tres zonas forman una especie de sandwich de zona p entre dos n.

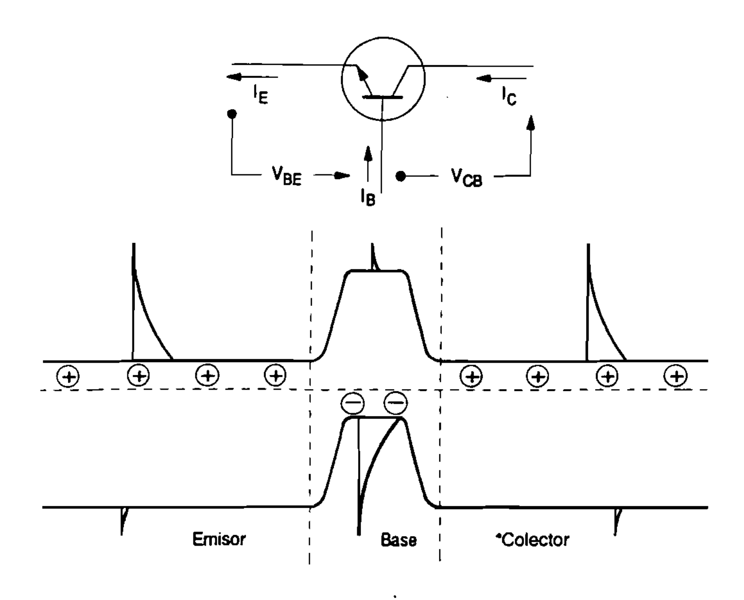


Fig. 5-17.

En la misma figura está representado el símbolo del transistor y las variables que definen su funcionamiento. Se han definido las variables de forma que sus valores sean positivos en funcionamiento normal del transistor. El transistor tiene tres terminales o **electrodos** que se llaman **emisor**, marcado por una flecha en el símbolo, **base**, dibujada con un trazo recto, y **colector**, tercer electrodo cuya misión veremos después. La **dirección** de la flecha del emisor indica que el transistor es npn. Existen también transistores pnp, y entonces la flecha se dirige hacia adentro.

El diagrama energético **sin polarizar** es el representado en la figura 5-17, y corresponde a dos diodos íntimamente unidos. En el diagrama puede verse el equilibrio en sentido vertical en las tres zonas y en sentido horizontal análogo al del diodo.

El transistor se **polariza** en **zona de trabajo** aplicando una polarización **a favor en el primer diodo**, V_{BE} de unas décimas de voltio (normalmente 0,5-0,7 V.) y una **una polarización inversa**

del segundo diodo, V_{CB} que puede variar entre cero y unos voltios (normalmente de 5 a 12 V.).

La figura 5-18 es el diagrama energético del transistor correctamente polarizado.

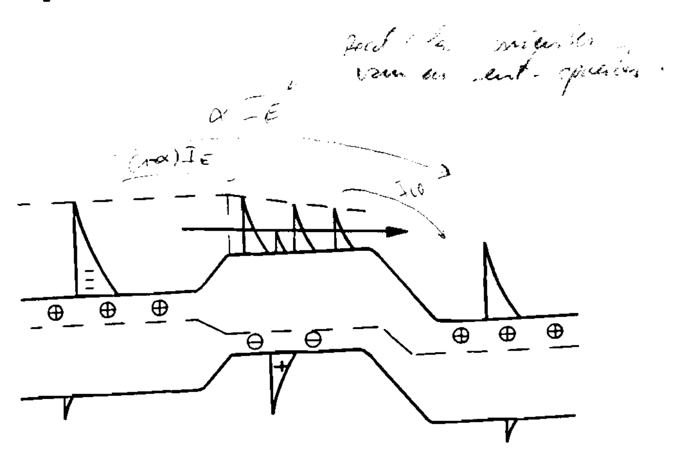


Fig. 5-18

Las ecuaciones del transistor se deducen con facilidad teniendo en cuenta la influencia de un diodo sobre el otro. El método que vamos a usar es una simplificación razonable del método seguido por Ebers y Moll. Proc. IRE, dic. 1.954. Este método se ha popularizado tanto que a las ecuaciones aquí obtenidas se les llama a veces de Ebers y Moll.

Si no existiese el segundo diodo, la polarización a favor del primero produciría una corriente de electrones de emisor a base de valor:

$$I_E=I_{E0}(e^{\frac{qV_{BE}}{KT}}-1)\approx I_{E0}e^{\frac{qV_{BE}}{KT}}$$

Si los diodos estuviesen muy separados, estos electrones, procedentes de las cargas libres mayoritarias del emisor, saldrían por la base. La polarización en contra del segundo diodo produce una pequeña corriente inversa de cargas minoritarias de la base al colector, de valor I_{co}, o sea, convencional de colector a base.

Sin embargo, la estrechez de la base, hace que, al difundirse las cargas mayoritarias del emisor en la base (ver fig.5-18) **la llenan** y aparecen frente a niveles vacíos del colector, al que pasan directamente en su mayor parte. Una pequeña parte sale por la base. Si llamamos α a la proporción de

cargas que saltan directamente al colector (α es próximo a 1, normalmente de 0,98 a 0,999) el colector recibirá la corriente de cargas negativas αI_E y la base $(1-\alpha)$ I_E. La corriente de colector será la suma de esta fracción de la de emisor y de la inversa de colector, ambas de **corriente convencional entrando en el colector**. Es decir, la corriente de colector vale:

$$I_C=\alpha I_E+I_{CO}$$

La corriente de base está, a su vez, compuesta por la fracción 1 - α de la de emisor (cargas negativas saliendo, corriente convencional positiva) y la corriente I_{CO} que sale. O sea:

$$I_B=(1-\alpha)\,I_E-I_{CO}$$

que también podría haberse obtenido restando I_C de I_E.

5.17.- Parámetros y características del transistor npn.

La figura 5-19 muestra el transistor npn y las variables que lo caracterizan. La única variación respecto al párrafo anterior es que se toma el emisor como origen de potenciales, con lo que la variable V_{CE} sustituye a la V_{CB} del párrafo anterior. Esta sustitución no tiene relevancia, ya que la tensión de colector no figura en las ecuaciones.

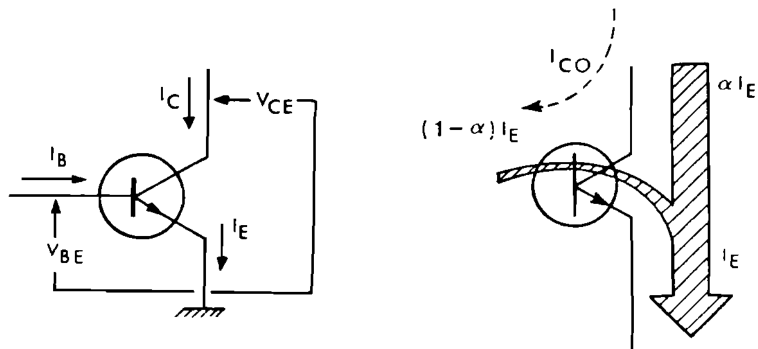


Fig. 5-19.

A la derecha de la misma figura se ve una imagen del funcionamiento del transistor. La fuerte corriente de emisor, procedente en su mayor parte del colector, es controlada por la tensión en la base. Una pequeña parte de la corriente entra por la base. Además, una pequeña corriente I_{CO} circula del colector a la base.

Las ecuaciones del transistor son :

$$I_E = I_{E0} e^{39 V_{BE}}$$
$$I_C = \alpha I_E + I_{CO}$$
$$I_B = (1-\alpha) I_E - I_{CO}$$

Hemos tomado q/KT = 39 V^{-1} , que es el valor 25ºC.

El transistor es un cuadripolo (activo y no lineal), cuyas **dos** ecuaciones son dos de las tres anteriormente citadas. De esas tres ecuaciones, sólo dos son realmente independientes ya que la tercera es consecuencia de las otras, por ser cero la suma de intensidades que entran en el transistor. Las variables de entrada son I_B y V_{BE} y las de salida I_C y V_C.

Los parámetros que figuran en las ecuaciones y que determinan el comportamiento del transistor son:

I_{E0} es la corriente inversa del diodo emisor-base, pero aquí es el parámetro que expresa la magnitud del control que V_{BE} ejerce sobre la corriente de emisor y también sobre la de colector, que es casi igual a la de emisor. En la práctica, no se maneja este parámetro, sino que suele conocerse una pareja de valores I_E, V_{BE} (a veces se toma I_C, V_{BE}, que,como veremos, es lo mismo).

I_{co} es la corriente inversa del diodo colector -base, y es casi siempre despreciable, especialmente frente a la corriente de colector, aunque no siempre frente a la corriente de base.

α es la relación de corriente dI_C / dI_E, o también I_C / I_E, despreciando I_{co}. El valor de α es muy próximo a la unidad, por lo que, en muchos casos puede tomarse $I_E = I_C$. A veces, en lugar del parámetro α se utiliza otro análogo llamado β, que se define como la relación dI_C / dI_B, **amplificación de corriente**, y cuyo valor, deducido fácilmente diferenciando las ecuaciones, será:

$$\beta = \frac{\alpha}{1-\alpha} \approx \frac{1}{1-\alpha}$$

de donde se deduce también la relación inversa:

$$\alpha = \frac{\beta}{1+\beta} \approx 1 - \frac{1}{\beta}$$

El parámetro β vale normalmente entre 50 y 300 y es más manejable que α.

En general, en la práctica, se suele despreciar I_{co}, con lo

que el comportamiento del transistor está suficientemente definido por los parámetros:

Una pareja de valores I_C, V_{BE}.

La amplificación de corriente, α o β.

Generalmente, los fabricantes dan estos parámetros y, además, recomiendan que se polariza en el punto de trabajo definido por la pareja de valores dados.

A continuación vamos a definir las **características del transistor**, que son las familias de curvas que definen su comportamiento. En la figura 5-19 vemos las cuatro variables externas que caracterizan al transistor, dos de salida, V_{CE}, en adelante V_C, e I_C y dos de entrada, V_{BE} e I_B. Como disponemos de dos ecuaciones, podemos eliminar una variable y quedarnos con una ecuación que relaciona las otras tres.

Así, si se elimina V_{BE}, se obtiene la relación:

$$I_B=(1-\alpha)I_E-I_{CO}=\frac{1-\alpha}{\alpha}(I_C-I_{CO})-I_{CO}, \text{ o sea}$$

$$I_C=\frac{\alpha}{1-\alpha}(I_B+I_{CO})+I_{CO}=\beta(I_B+I_{CO})+I_{CO}$$

Sorprende que en esta relación **no figure V_C**, irregularidad matemática típica de la cualidad separadora del transistor, como veremos.

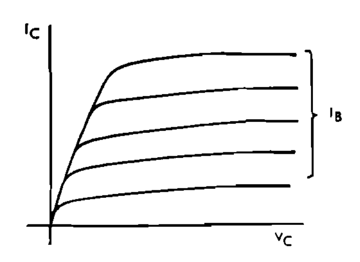


Fig. 5-20.

La ecuación anterior permite ahora dibujar las curvas de dependencia de I_C en función de V_C tomando I_B como parámetro. Debido a la forma tan lineal de la ecuación, las curvas teóricas serían rectas horizontales (ya que I_C no depende de V_C) regularmente espaciadas. Las **características reales**, tomando I_B como parámetro, están representadas en la izquierda de la figura 5-20. La diferencia entre las curvas teóricas y las prácticas puede resumirse en los dos puntos siguientes:

Las curvas se inician en el origen y ascienden rápidamente hasta alcanzar un segundo tramo casi horizontal. Esto es debido

a que en la zona de la banda izquierda no se cumplen las hipótesis con las que hemos planteado nuestras ecuaciones, en particular, la polarización del diodo base-colector no es inversa hasta que V_C alcance cierto valor, sobrepasando a V_{BE}. En esa zona, llamada **de saturación**, el transistor se comporta como una resistencia.

Las curvas no son horizontales en su segundo tramo, sino que muestran una ligera pendiente que expresa la influencia de V_C sobre I_C y que no aparece en las ecuaciones debido a las simplificaciones realizadas. Esta influencia es una **admitancia de salida**, llamada $\mathbf{h_{22}}$ porque coincide con este parámetro híbrido. Su valor suele ser muy pequeño, menor de 10^{-6} mhos y se desprecia casi siempre en la práctica. En la figura se ha representado muy exagerado para ponerlo de manifiesto, pero en las curvas reales no sería perceptible.

Eliminando ahora I_B (no usando su ecuación), se obtiene la otra familia de características teóricas, de I_C en función de V_C (que no aparece) y tomando como parámetro V_{BE}.

$$I_C=\alpha I_E+I_{CO}=\alpha I_{E0}e^{\frac{q}{KT}V_{BE}}+I_{CO}$$

A esta ecuación le corresponde una familia de rectas horizontales (porque no aparece la dependencia de V_C) **espaciadas exponencialmente**, con un fuerte crecimiento en función del parámetro. La derecha de la figura 5-21 representa las **características reales**, que se diferencian de las teóricas en los dos mismos puntos que antes.

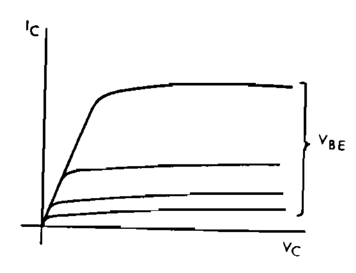


Fig. 5-21.

5.18.- Transistores bipolares npn y pnp.

Podemos considerar dos tipos de transistores bipolares, formados por dos uniones próximas entre sí:

En primer lugar el transistor **npn** que hemos visto hasta ahora. Su diagrama energético sin y con polarización es el de las figuras 5-17 y 5-18. El funcionamiento del transistor está totalmente ligado a las **cargas libres negativas**, tal y como lo

hemos explicado. Las cargas libres positivas, mayoritarias en la base, no tienen movimientos relevantes para el transistor, como es fácil comprobar a la vista del segundo diagrama, ya que las cargas mayoritarias de la base, que pasan al emisor, vacían ésta enseguida (no hay que olvidar el carácter geométrico del funcionamiento) y no dan lugar a una corriente apreciable.

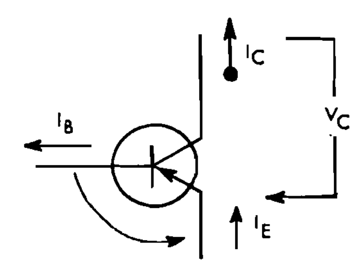


Fig. 5-22.

En segundo lugar el transistor **pnp**, que precedió históricamente al anterior (los primeros transistores eran pnp de germanio) y que funciona de forma análoga por difusión de **cargas libres positivas** en la base y paso posterior al colector. La figura 5-22 representa un transistor pnp . Fijando adecuadamente los signos convencionales de las variables, como se ha hecho en la figura, **las ecuaciones y características son las mismas del párrafo anterior**. Como se observa, todos los signos convencionales de tensiones y corrientes son los contrarios a los del npn.

Con mucha diferencia, el más usado es el **transistor npn de silicio**, al que en adelante llamaremos simplemente **transistor**. El transistor pnp ha quedado relegado a ser complementario del anterior o en algunos circuitos en que la alimentación negativa es conveniente.

5.19.- Ecuaciones linealizadas para pequeñas señales.

Las ecuaciones del transistor son (recordemos que sólo dos son independientes):

$$I_E = I_{E0} e^{39 V_{BE}}$$
$$I_C = \alpha I_C + I_{CO}$$
$$I_B = (1-\alpha) I_E - I_{CO}$$

Diferenciando estas ecuaciones, obtenemos :

$$dI_E = 39\, I_{E0} e^{39 V_{BE}} dV_{BE} = 39\, I_E\; dV_{BE}$$
$$dI_C = \alpha\, dI_E$$
$$dI_B = (1-\alpha)\, dI_E$$

Si ahora sustituimos las diferenciales por incrementos finitos pequeños y llamamos a estos incrementos, que constituyen las señales con minúsculas, obtenemos las ecuaciones **linealizadas** para un **punto de trabajo** (el punto en que hemos diferenciado):

$$i_E = 39\, I_E v_{BE}$$
$$i_C = \alpha\, i_E$$
$$i_B = (1-\alpha)\, i_E$$

Estas ecuaciones son lineales, despreciando la variación de I_E con la señal, es decir , los infinitésimos de segundo orden. Hay que hacer notar que la variable I_E, única que queda del **punto de trabajo**, permanece como un parámetro de las ecuaciones lineales. Es decir que el punto de trabajo influye en este parámetro del cuadripolo lineal, para señales.

Considerando ahora al transistor como un cuadripolo para señales, sus ecuaciones son las linealizadas (recuérdese que sólo dos son independientes). Las variables de entrada del cuadripolo son v_{BE}, i_B y las de salida v_C, i_C. Las ecuaciones del cuadripolo son (eliminando i_E):

$$v_{BE} = \frac{1}{39\, I_E (1-\alpha)}$$
$$i_C = \frac{\alpha}{1-\alpha}\, i_B = \beta\, i_B$$

Estas ecuaciones tienen la forma de las de parámetros híbridos, que eran (teniendo en cuenta la denominación que hemos dado a las variables de entrada y salida en el caso del transistor:

$$v_{BE} = h_{11} i_B + h_{12} v_C$$
$$i_C = h_{21} i_B + h_{22} v_C$$

Ahora podemos identificar los parámetros híbridos, cuyo valor será :

$$h_{11} = \frac{1}{39\, I_E (1-\alpha)} \approx \frac{\beta}{39\, I_E} \quad h_{21} = \beta \quad h_{12} = h_{22} = 0$$

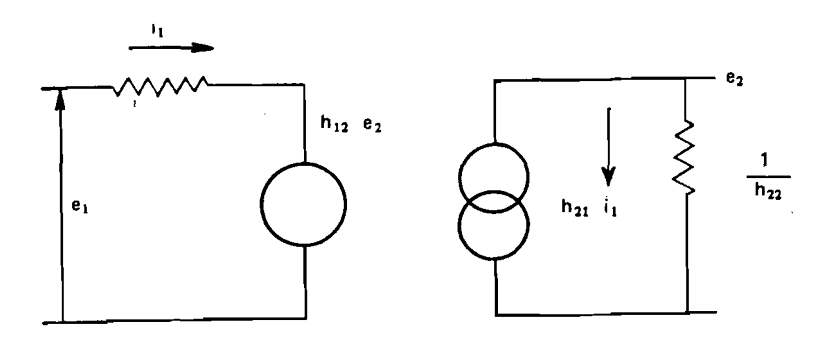


Fig. 5-23.

La figura 5-23 es el **circuito equivalente** del transistor, poniendo de manifiesto el valor de los parámetros híbridos. En nuestro caso, el generador de tensión $h_{12}v_C$ es nulo. El valor de $1/h_{22}$, que no figura en las ecuaciones, puede dejarse como corrección práctica.

CAPITULO 6

SEMICONDUCTORES ESPECIALES

6.1.- Introducción.

En el capítulo anterior hemos visto los fundamentos de semiconductores, iniciándonos en las propiedades físicas y los primeros dispositivos. En este capítulo comenzaremos por explicar brevemente las técnicas de fabricación y su evolución hasta las técnicas de producción de circuitos integrados. También describiremos otros dispositivos semiconductores, como transistores MOSFET, diodos Zener, etc... limitándonos a algunos tipos importantes pero que sirven de modelo para comprender los dispositivos semiconductores que surgen en el mercado con cierta frecuencia, algunos para desaparecer al cabo de poco tiempo, sustituidos o superados por otros.

6.1.- Semiconductores de aleación o soldadura y crecimiento.

Los primeros transistores se fabricaban por soldadura o aleación (alloy) **uniendo** fragmentos de silicio de tipo p o n para formar diodos y transistores. Han sido los primeros diodos y transistores, si se exceptúan los de "contacto" utilizados con anterioridad, cuyas partes se unían solo por presión. La aleación es una soldadura autógena formándose una zona de transición gradual en la unión, aunque aparece una discontinuidad de la red cristalina. El método se aplica ya poco, al sustituirse por otros más perfeccionados, pero sigue siendo práctico para algunos transistores de gran potencia que no pueden fabricarse por otros procedimientos.

Otros transistores se fabrican cambiando el dopado de la fase líquida durante la formación o "crecimiento" del cristal de silicio. Con ello se obtiene transistores (grown) con una

transición más perfecta entre las zonas p y n , ya que el **substrato de silicio** forma casi un monocristal, aunque en la práctica también se forman dislocaciones con el cambio de dopado.

La formación del monocristal se hace comenzando con un pequeño fragmento catalizador y extrayéndolo lentamente del líquido, manteniendo en todo momento el contacto sólido-líquido.

La figura 6-1 representa estos dos tipos de transistores.

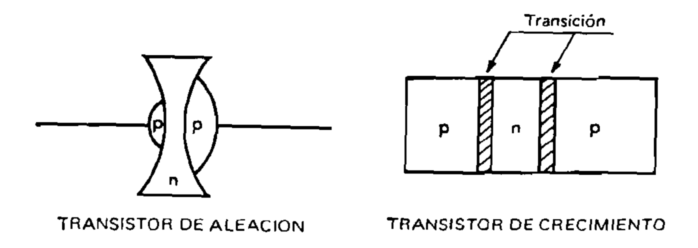


Fig. 6.1

6.2.- Semiconductores de difusión, semiconductores planar epitaxiales.

La gran mayoría de los semiconductores actuales se fabrican por **difusión planar epitaxial.**

La técnica de difusión planar consiste en someter una cara (de ahí el nombre de **planar**) de una "galleta" u **oblea** de silicio puro a la acción de un gas a presión y temperatura convenientes. Las partículas del gas se difunden (penetran) en el seno del sólido en las proximidades de la superficie, introduciendo las impurezas adecuadas. La difusión se llama **epitaxial** si el gas que se difunde contiene también ión silicio, con lo que se forma un contacto de fase sólida y gaseosa que favorece el intercambio de átomos introduciendo además pequeñas proporciones de impurezas que realizan el **dopado.** La impureza está presente en el gas en forma de iones de átomos donantes o aceptantes.

Todas las operaciones planares se hacen a través de **ventanas** que se abren en la capa de óxido de silicio que cubre y protege al silicio puro. Las operaciones se hacen sucesivamente y las ventanas se cierran oxidando el silicio y se abren otra vez reduciendo la capa de óxido (sílice).

Las operaciones de dibujo y tratamiento de las ventanas se hacen utilizando técnicas de **fotohuecograbado.**

El huecograbado es un procedimiento clásico de fabricación de bajorelieves, por ejemplo, para dibujar figuras en placas de cobre. En este caso, se cubría la placa de cobre con una sustancia inerte, como cera o parafina, formando un negativo de la figura a dibujar. A continuación, se atacaba el metal con un ácido corrosivo, por ejemplo FlH, que dibujaba la figura. Luego

se lavaba cuidadosamente y se quitaba por calor la capa de cera o parafina. La figura era el equivalente de nuestra ventana actual.

Además del huecograbado, se pueden usar técnicas fotográficas para realizar el depósito negativo de la ventana, lo que constituye un fotohuecograbado.

Veamos los pasos a seguir para fabricar un transistor con técnica planar epitaxial y métodos de fotohuecograbado (fig. 6-2)

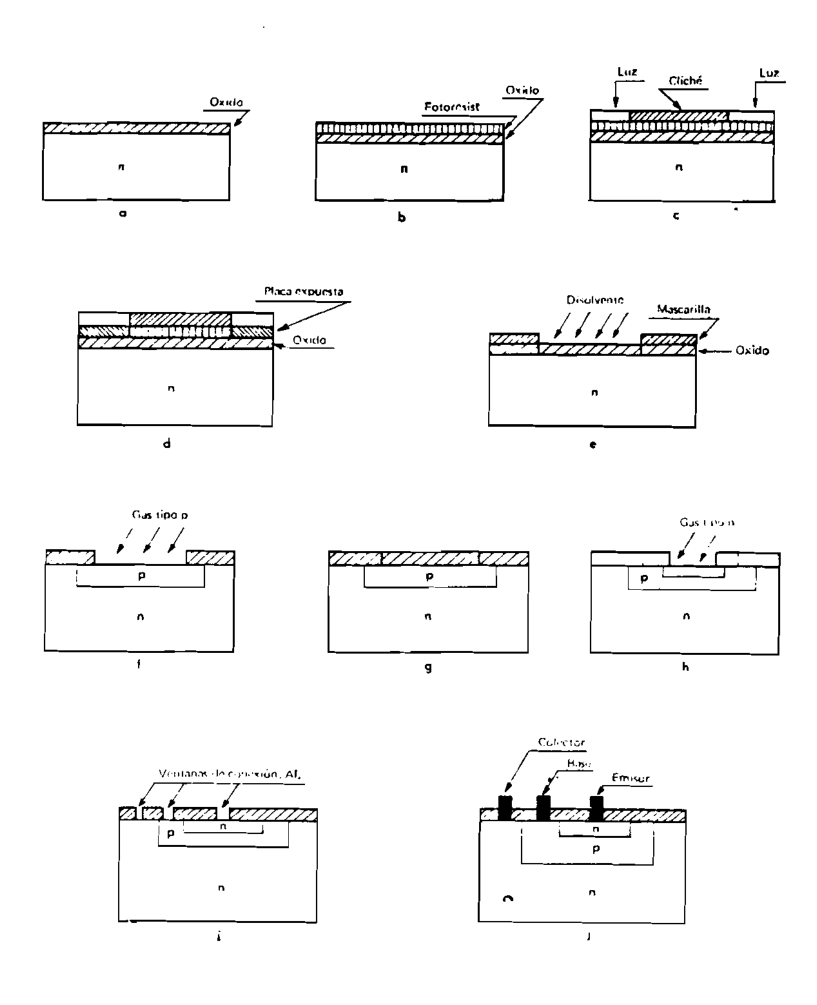


Fig. 6-2.

a) Se parte de un **substrato** de silicio de tipo n, cubierto de una capa de protección de óxido de silicio (SiO_2), (fig. 6-2a).

b) Se cubre la capa de óxido de un material sensible a la luz, para, posteriormente "dibujar" la ventana. Por ejemplo, "fotoresist" (Fig. 6-2b).

c) Se coloca encima un cliché negativo de las zonas a abrir (Fig. 6-2, c y d). Se expone a luz ultravioleta, a la que es sensible el fotoresist.

d) Se aplica un disolvente que elimine la sustancia de las zonas no expuestas a la luz.(A veces hay que "fijarlo" antes). A lo que queda se le llama **mascarilla**. La mascarilla deja una **ventana**, a través de la cual se puede ya operar (fig.6-2e).

e) Se aplica un reductor que elimine el óxido en la ventana o superficie libre de mascarilla.

f) Se expone al silicio libre de protección a la acción de un gas que contenga ión silicio (Cl_4 Si) e impurezas de boro (B_2H_6) que forma la impureza de tipo p. Este gas, a alta presión y alta temperatura se difunde **epitaxialmente** formando una zona de tipo p en el interior del **substrato** de silicio. Esta zona no estará tan claramente delimitada como aparece en la figura 6-2f,y forma con el substrato un diodo pn.

g) Se vuelve a oxidar el silicio (fig. 6-2g).

h) Se repiten las operaciones b) a e) para abrir una ventana menor que la anterior. Al final se expone al silicio a un gas con una impureza de tipo n, por ejemplo PH_3 (fig. 6-2h).

i) Se oxida de nuevo y se repite todo el proceso b) a e) para abrir unas pequeñas ventanas de conexión (fig. 6-2i).

j) Se deposita aluminio electrolítico en las ventanas y se hacen las conexiones exteriores. El transistor queda así terminado y protegido bajo una capa de silicio. (fig.9-2j).

6.3.- Circuitos integrados monolíticos.

La fabricación planar se presta especialmente a la fabricación en serie, ya que pueden fabricarse un gran número de transistores (u otros dispositivos) sobre una pastilla de silicio, por difusión planar epitaxial desde un lado y luego cortar la pastilla obteniendo una gran cantidad de componentes de una misma pastilla.

En seguida surge la idea de que, para un circuito que necesite varios transistores, no es necesario cortar la pastilla, sino que basta con realizar las conexiones exteriores sobre la misma pastilla, formando así un circuito **integrado** sobre una sola

pastilla, por lo que se llama **monolítico**.

El substrato se polariza, como veremos a continuación, aislando así todos los componentes entre sí. Si se trata de un substrato de tipo p, por ejemplo, se polariza el substrato al potencial más negativo presente en el circuito, como se ve en la figura 6-3. Todas las uniones con el resto de los componentes están así sometidas a una polarización inversa y aparecen "aisladas" ya que ese tipo de polarización evita la conducción. Como se ve en la figura, en el circuito equivalente aparecen diodos polarizados en contra, que no influyen en el funcionamiento del circuito.

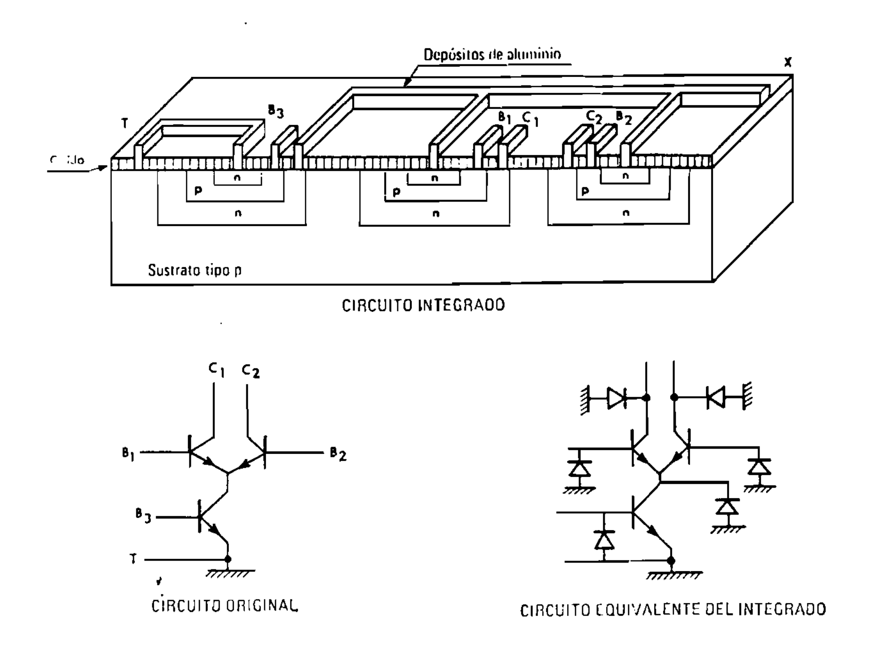


Fig. 6-3.

Las conexiones exteriores entre los componentes se hacen con un depósito de aluminio electrolítico, como se ve en la figura, lo que equivale a un circuito "impreso" que se hace con las mismas técnicas de huecograbado que hemos visto para la fabricación del circuito integrado, sólo que aquí se deposita el aluminio a través de la máscara en lugar de dopar con un gas a presión.

El circuito integrado (**IC**, integrated circuit) se construye por estratos alternados p y n, pero la forma y profundidad es diferente en diferentes puntos, debido a que el sistema de ventanas tiene una gran flexibilidad. Lo más típico es la fabricación de transistores, que también pueden usarse como **diodos**. Otros componentes se fabrican o "integran" como sigue:

a) Resistencias.

Se fabrican con un fragmento convenientemente dopado para obtener la conductividad que se desee y en forma de canal para aumentar la resistencia. Si, como se ve en la figura 6-4, se usa el mismo tipo de dopado que el substrato, se puede aislar con un "lecho" de material de signo contrario, lo que forma un aislamiento en doble diodo como el representado en la figura. Obsérvese que la distribución pnp es parecida a la de un transistor, pero no tiene el mismo comportamiento porque la base no es estrecha y los dos diodos **no están acoplados entre sí.** Es posible obtener una gama de valores de unos ohmios a unos kiloohmios, con las pequeñas dimensiones de los circuitos integrados.

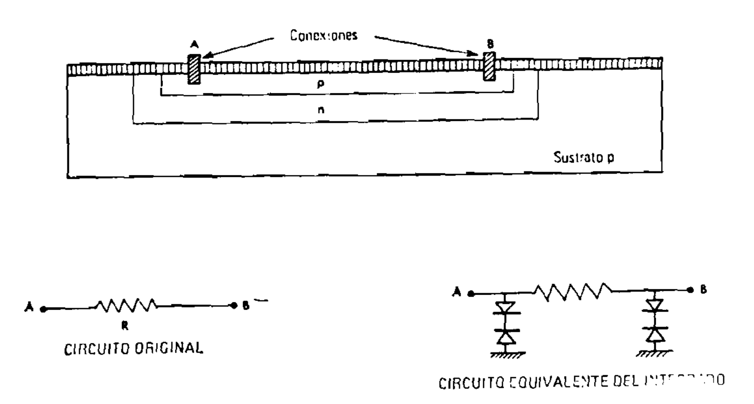


Fig. 6-4.

b) Condensadores.

Se utiliza la capacidad de una unión pn polarizada en contra o se utiliza un electrodo semiconductor y otro de aluminio, con SiO_2 como dieléctrico, como se ve en la figura 9-5. No es posible, debido al pequeño tamaño de los circuitos integrados, obtener valores altos de capacidad, sólo de unos picofaradios.

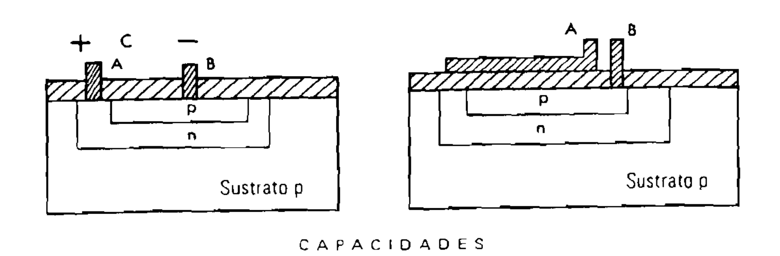


Fig. 6-5.

c) Autoinducciones.

Es muy difícil obtener valores apreciables de la autoinducción, en recorridos tan cortos. Se puede fabricar como

se ve en la figura 9-6, como "película fina" de aluminio depositado, pero se evita en lo posible.

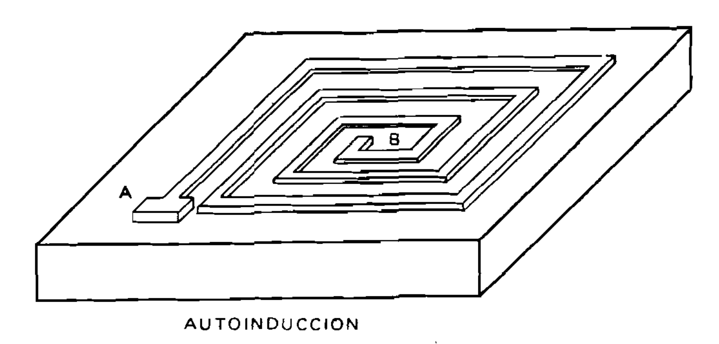


Fig. 6-6.

6.4.- Otras técnicas de miniaturización.

Las técnicas de huecograbado y fotohuecograbado citadas se aplicaron ya antes de los circuitos integrados en los llamados **circuitos impresos**.

Los circuitos impresos son depósitos de material conductor, aluminio o **cobre**, sobre una tarjeta de material aislante, de cartón endurecido o resina sintética, para formar con su dibujo todas las conexiones del circuito, sin componentes. Los componentes se **sueldan** al depósito por medios más o menos automatizados, constituyendo así el resto del circuito. Las conexiones exteriores del circuito, como alimentación, entradas y salidas, se "llevan" a un borde y se unen a un **conector** de contactos múltiples, o, como se ha indicado en la figura, se utilizan directamente como contactos de conexión exterior en un conector acoplado. Las tarjetas que forman parte de un sistema se conectan lado a lado como los libros en una librería. La figura 6-7 representa una tarjeta impresa correspondiente al circuito allí representado.

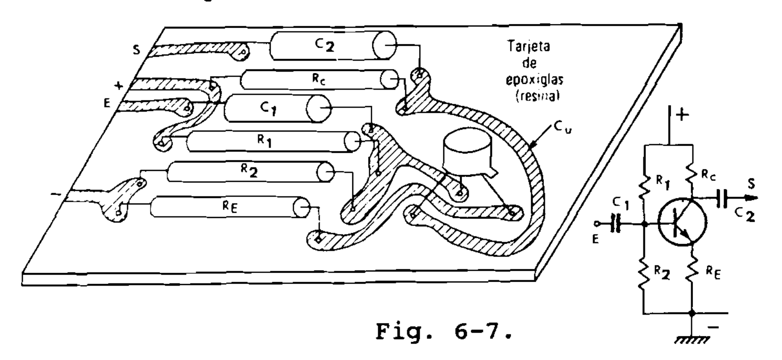


Fig. 6-7.

Las tarjetas impresas han sobrevivido a los circuitos integrados y usan a estos circuitos como componentes, realizando todas las conexiones de los IC entre ellos y también agregando

otros componentes.

Actualmente se usan las tarjetas impresas por los dos lados, constituyendo así cada tarjeta un conjunto de dos circuitos impresos que se unen entre sí por conductores canalizados a través de pequeños taladros. También pueden agruparse varias tarjetas formando un paquete y con conexiones perpendiculares a través de taladros, lo que constituye la técnica **multilayer** (multicapa).

Otras técnicas de miniaturización que conviene citar son las de **película fina** (thin film) y **película gruesa** (thick film). La primera consiste en depositar una fina capa sobre una base cerámica y construir con ella componentes como resistencias (con material resistivo), líneas de transmisión o autoinducciones. En "capa gruesa" se pegan además otras componentes como cristales de cuarzo, etc... Cuando estas técnicas se usan tomando como base un circuito monolítico, se obtienen circuitos integrados llamados **híbridos**.

La utilización de circuitos monolíticos con gran cantidad de componentes se llama **MSI**, medium scale integration o **LSI**, large scale integration. Con ellos se consigue integrar sistemas enteros, como un receptor, una unidad de cálculo, etc... en un solo circuito.

6.5.- Diodo Zener.

Además de los diodos y transistores han aparecido otros dispositivos semiconductores basados en la Física del Estado Sólido y de los que mencionaremos los más conocidos. El primero de ellos es el diodo Zener, basado en el **efecto Zener**.

El efecto Zener se presenta en los diodos con polarización inversa en la forma siguiente:

Cuando la tensión inversa llega a un valor determinado, los bordes de la banda de conducción del lado n y de la banda de valencia del lado p coinciden en nivel energético como se ve en la figura 6-8.

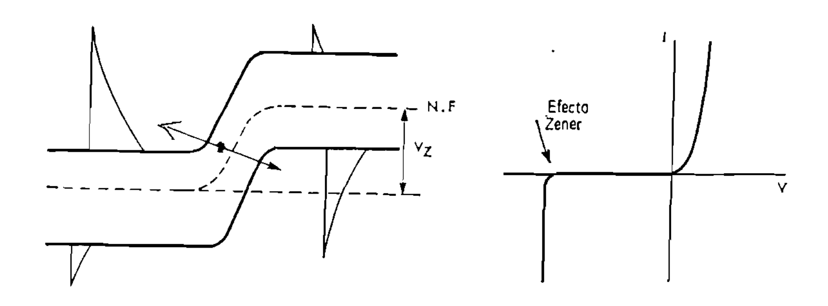


Fig. 6-8.

La disociación de cargas ocurre ahora entre los dos lados, saltando los electrones de la banda de valencia del lado p a la banda de conducción del lado n por **disociación** y conduciendo fácilmente a través de las dos zonas mayoritarias. Este estrechamiento entre ambas zonas forma un **túnel** de naturaleza parecida al que veremos después. En la característica del diodo aparece una conducción brusca, como se ve en la figura, que aparece a la **tensión Zener**.

Aunque todos los diodos presentan el efecto Zener, en los diodos normales aparece a tensiones diferentes en diferentes puntos del diodo, y la curva no es tan brusca como la representada en la figura.

El diodo Zener se polariza con tensión inversa y conduce a partir de la tensión característica E_z, como se ve en la figura 6-9. El símbolo es un diodo con una Z. La característica sube a partir de E_z con una pendiente no infinita, mostrando una **resistencia diferencial r_d**. La ecuación del diodo Zener es :

$$E=E_z+r_dI$$

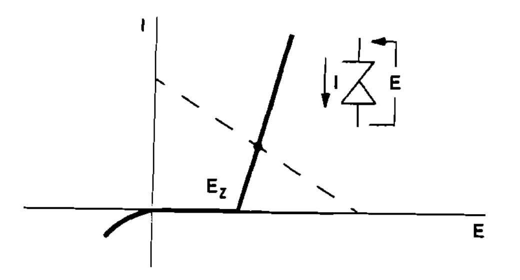



Fig.6-9.

La ecuación del diodo Zener sólo es válida para I>0, o sea $E > E_z$. Si la tensión es inferior a E_z, I=0, el diodo no conduce y se porta como un diodo normal. Si E < 0, conduce en el sentido normal de un diodo. Con una recta de carga como la representada en la figura, en el Zener aparece una tensión casi igual a su valor característico E_z, "estabilizando" así la salida, como veremos con más detalle en un capítulo posterior.

Los diodos Zener se fabrican en una gama muy extensa de tensiones Zener (variando el dopado), desde unos voltios hasta algunas decenas de voltios. Su utilización como **estabilizador de tensión** será estudiada más adelante.

6.6.- Diodo Esaki o túnel.

En materiales fuertemente dopados los bordes de la banda de conducción del lado n y banda de valencia del lado p están situados al mismo nivel incluso sin polarización, con lo que con una polarización directa pueden pasar directamente los electrones de la zona n a la zona p a ocupar los huecos mayoritarios del mismo nivel a través del **túnel** o estrecha distancia entre ambas zonas (ver fig. 6-10). Se produce una fuerte corriente eléctrica por **recombinación** de cargas positivas y negativas, lo que se llama **efecto túnel o Esaki**.

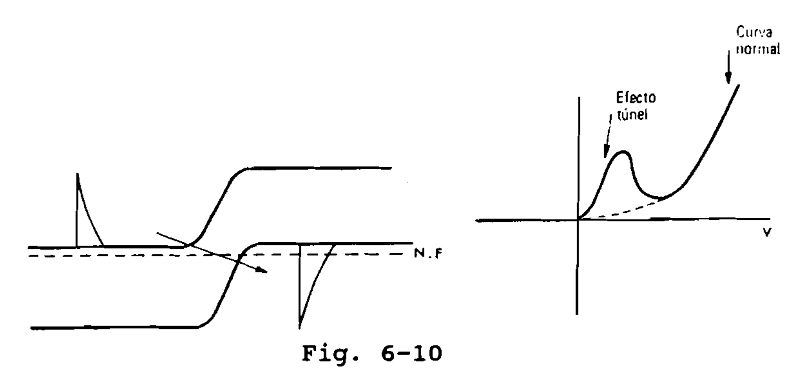


Fig. 6-10

Al aumentar la polarización directa, este efecto desaparece y el diodo presenta un comportamiento normal, conduciendo por difusión de cargas mayoritarias en la zona minoritaria. Por eso en la característica, dibujada en la figura, aparece una prominencia inicial debida al efecto Esaki, que luego desaparece.

Las aplicaciones del diodo túnel están basadas en que si se traza una recta de carga en la figura, corta a la característica en más de un punto, ocurriendo inestabilidades debidas al cambio de punto de trabajo (salto de uno a otro), que dan lugar a oscilaciones de muy alta frecuencia. También puede utilizarse la gran pendiente inicial de la característica.

6.7.- Transistores de efecto de campo (FET).

Si se polariza en sentido inverso una unión pn, como se ve en la figura 6-11, se obtiene una zona de **deplexión** o empobrecimiento (depletion) con ausencia de cargas libres, que son rechazadas por la polarización negativa del electrodo p. Bajo este principio se puede utilizar el **FET (field effect transistor) o junction FET**, así llamado porque, en realidad, es un dispositivo con una unión pn. La zona n constituye un **canal** entre dos electrodos que se denominan **fuente (source) y sumidero (drain)**. Cuando no hay polarización las cargas mayoritarias (en este caso negativas) permiten una conducción considerable entre S y D. Cuando se polariza el tercer electrodo o **puerta (gate)**, se forma la zona de **deplexión** que **estrangula** el canal y la conducción cesa.

La característica o dependencia de la corriente I_D en función de V_{GS} está representada en la figura 6-12. La pendiente de la curva permite **controlar** la intensidad del canal con la

tensión en la puerta y así **amplificar**. Además, la impedancia de entrada es muy elevada, debido a la polarización inversa, lo que nos permite aplicar este tipo de dispositivo en puntos donde no se puede perturbar al sistema con una intensidad exterior.

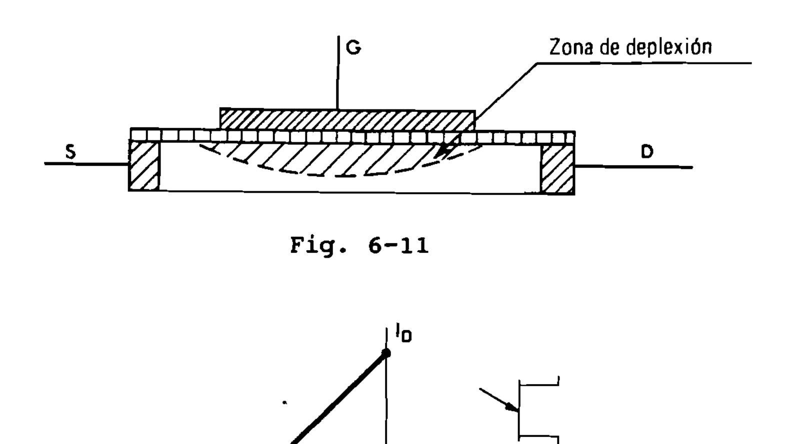


Fig. 6-11

Fig. 6-12

La figura 6-13 representa las características de salida del dispositivo, es decir I_D en función de V_{DS} y V_{GS} como parámetro. Las curvas son análogas a las de los transistores bipolares. La tensión de saturación de aquellos está sustituida aquí por la **tensión de estrangulamiento (pinch off)**, que es la tensión a la cual se cierra el canal . Por debajo de esa tensión el ancho del canal depende de V_{DS} y la intensidad crece con la tensión. A partir del valor de estrangulamiento, la intensidad crece muy poco con la tensión, como ocurría en los transistores bipolares. Los transistores FET sustituyen a los bipolares sólo en aplicaciones que exigen una muy alta impedancia de entrada.

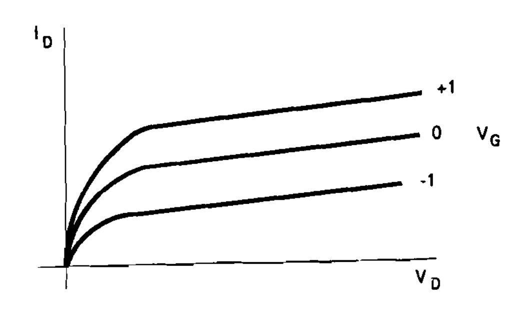


Fig. 6-13.

6.8.- Rectificador controlado (SCR).

El **SCR (silicon controlled rectifier)** es un dispositivo de **cuatro** capas de semiconductor dopado alternativamente, por lo que también se le llama **four layers device**. Se trata de un **conmutador**, que sólo funciona en dos estados puntuales, "conduce" o "no conduce". Es un dispositivo **biestable**.

La figura 6-14 representa la construcción del dispositivo, el símbolo que lo representa y el circuito equivalente que sirve para comprender su funcionamiento. El dispositivo tiene tres electrodos, que se llaman **ánodo**, **cátodo** y **disparador (trigger)**. Los dos primeros son equivalentes a sus homónimos de un diodo y el tercero es el electrodo de control. A la vista de la figura de la izquierda, se ve fácilmente que las tres capas inferiores forman un transistor npn y las tres capas superiores, un transistor pnp. La base del npn (segunda capa desde abajo) es el colector del pnp, y el colector del npn es la base del pnp. Por eso resulta válido el circuito equivalente de la derecha. Polarizaremos el ánodo con una tensión positiva.

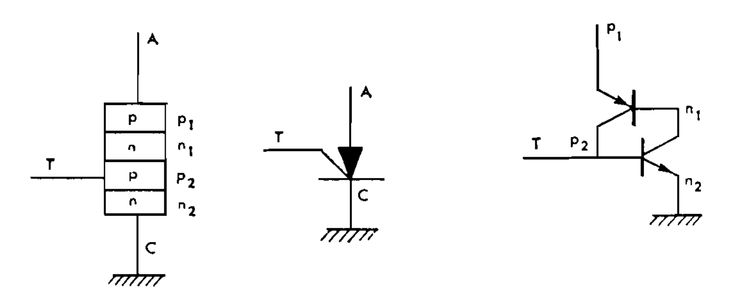


Fig. 6-14.

El primer "estado" es el de **conducción**. Cada transistor recibe su corriente de base del colector del otro transistor, por lo que ambos están fuertemente polarizados conduciendo en saturación. La corriente que "entra" por p_1, en sentido convencional, puede considerarse dividida en dos ramas iguales (esquema de la derecha), una que "salta" a p_2 y es la de colector del pnp y de base del npn y otra que sale por la base n_1 del pnp y "salta" al emisor n_2 del npn. Ambos transistores tienen corrientes de base iguales a las de colector y están en saturación. Esta división es puramente teórica, ya que, en realidad, las dos mitades atraviesan juntas las cuatro capas. El dispositivo conduce como un diodo con polarización directa.

El segundo "estado" es el de **no conducción**. Cada transistor no tiene corriente de base y cierra al otro al no suministrarle corriente de base. La polarización es baja (próxima a cero) en p_2 y alta en n_1. El diodo no conduce. En la figura de la izquierda, las uniones pn superior e inferior tienen muy poca caída de potencial, y toda la tensión exterior está aplicada en contra de la unión np intermedia. El dispositivo no conduce, como un diodo (el de las capas centrales) polarizado en contra.

Ambos estados son estables y cualquier situación intermedia es inestable porque los dos transistores del circuito equivalente

están unidos con una amplificación en circuito cerrado positiva y alta, lo que, como veremos más adelante, produce una fuerte inestabilidad. Si se polariza en sentido contrario, el dispositivo no conduce tampoco, por estar los dos diodos polarizados en contra.

El paso de **no conducción a conducción** se hace aplicando un corto **impulso positivo** al electrodo de disparo o **trigger**. Esto obliga a conducir al transistor npn y éste hace conducir al otro, permaneciendo ambos en estado de conducción después del trigger. El impulso puede ser muy corto, del orden de microsegundos.

El paso de **conducción a no conducción** es más difícil. No puede hacerse a través del trigger, como parecería a primera vista, ya que la corriente de base es muy alta y habría que absorberla toda durante un instante, lo que haría al dispositivo inútil, al tener que disponer de gran potencia de control. El único procedimiento de "corte" es quitar la polarización del ánodo durante un breve intervalo de tiempo. En realidad éste es el gran inconveniente de los SCR, que necesitan un **circuito de conmutación**, que anule durante un corto tiempo la tensión de ánodo. A la derecha de la figura 6-14 se ve la característica del SCR.

A pesar del inconveniente de la "conmutación", el SCR tiene interesantes aplicaciones en fuentes de alimentación conmutadas y en circuitos de control de gran potencia, como veremos más adelante.

6.9.- Transistores MOSFET.

Los transistores **MOSFET** (Metal Oxide Semiconductor Field Effect Transistor) son dispositivos de **canal**, análogos a los FET ya descritos, pero con la puerta aislada por medio del óxido de silicio (SiO_2) o nitruro de silicio (Si_3 N_4, lo que intensifica aun más el efecto de la alta impedancia de entrada.

Según el tipo de semiconductor que forma el canal o la forma de controlarlo, existen cuatro clases de MOSFET, que pueden ser de canal n o p y de "enriquecimiento" (enhacement) o de "empobrecimiento" (deplexion). Describiremos los cuatro, aunque los más usados son los de enriquecimiento de canal n y, mucho menos y generalmente como complementarios, los de enriquecimiento de canal p.

Aunque se pueden usar los MOSFET de forma análoga a los transistores bipolares, no son mejores que ellos en aplicaciones analógicas, excepto en algunos casos especiales, pero sí en las aplicaciones digitales, donde los sustituyen con ventaja.

6.10.- MOSFET de enriquecimiento canal n.

Está constituido, como se indica en la figura 6-15 por dos conexiones a zonas de dopado n sobre un sustrato de tipo p. Estos dos electrodos, llamados **S**, source (**fuente**) y **D**, drain (**drenaje o sumidero**) van a ser los extremos del canal. Se puede ver que

estos electrodos son intercambiables. El canal con los dos electrodos forma dos diodos conectados en oposición (o un transistor npn con la base desconectada) y no conduce normalmente, aunque se polarice el drenaje positivamente, como se hace en funcionamiento. Encima del "canal" todavía inexistente se coloca una **puerta** o **graduador** (gate) aislada a través del óxido de silicio, como se ve en la figura. Normalmente el "canal" es un paso estrecho entre los dos electrodos, como también se aprecia en la construcción indicada.

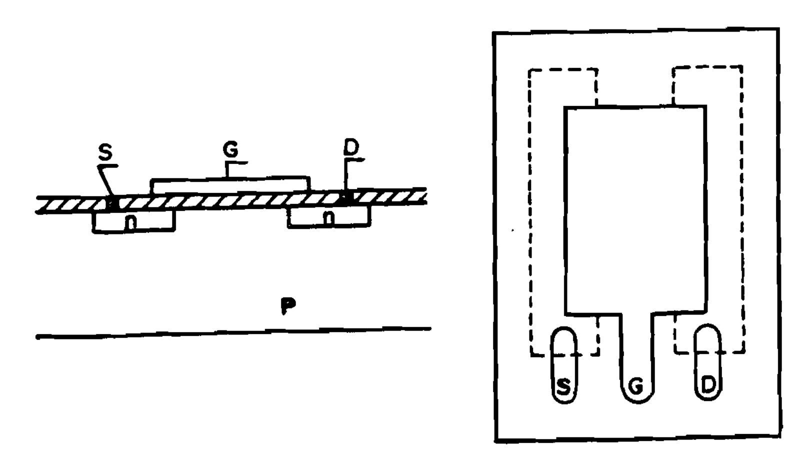


Fig. 6-15.

Cuando se conecta la puerta a una polarización positiva, se atraen las cargas libres negativas minoritarias de la zona p, como se ve en la figura 6-16. Este exceso de cargas libres negativas produce una zona de **enriquecimiento** (enhacement) con una **inversión** de la polaridad de las cargas libres y forma un canal de conducción. Es un efecto de campo, debido al campo eléctrico producido por la puerta.

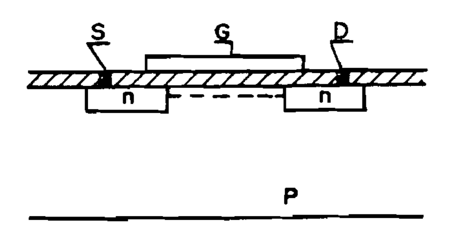


Fig. 6-16.

La figura 6-17 representa la aparición y aumento de la corriente de drenaje en función de la tensión de puerta (a polarización de drenaje constante) y es lo que caracteriza al efecto de **control** de la puerta. La figura 6-18 representa las características del transistor, análogas a las de un transistor bipolar, si bien el efecto de la tensión de drenaje se hace notar

en una zona mayor que la de saturación de los transistores bipolares, debido a que la tensión positiva de drenaje también contribuye a la formación del canal, de forma parecida al efecto de estrangulamiento de los FET.

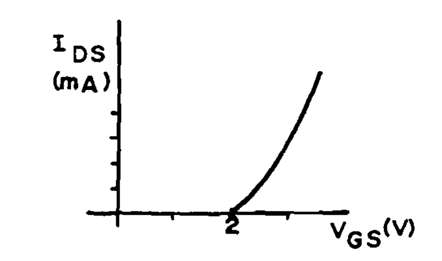


Fig. 6-17.

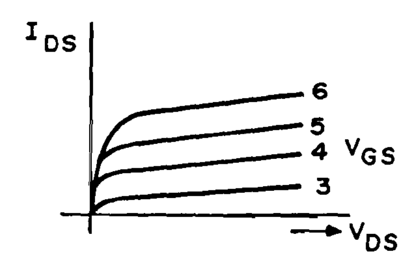


Fig. 6-18.

El transistor MOSFET de canal n de enriquecimiento es el más análogo a los transistores bipolares npn, hasta el punto de que puede usarse con los mismos circuitos. La deducción matemática de las características no es interesante, ya que la dependencia de la construcción y geometría del transistor hace que se alejen bastante de cualquier aproximación teórica, que sólo sería válida en algunos casos particulares.

6.11.- MOSFET de enriquecimiento de canal p.

En este caso el sustrato es de tipo n y el canal se forma entre dos electrodos de tipo p, como se ve en la figura 6-19.

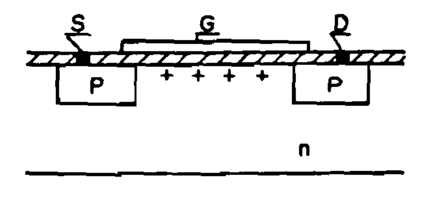


Fig. 6-19.

El funcionamiento es exactamente igual al anterior,

invirtiendo todos los signos. El drenaje se polariza negativamente y el dispositivo no conduce sin tensión en la puerta. Al aplicar una tensión negativa a la puerta se forma un canal p y se controla la conductividad con la tensión de puerta. Las características son las de la figura 6-20.

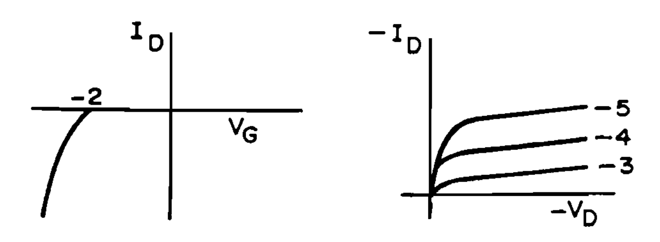


Fig. 6-20.

6.12.- MOSFET de empobrecimiento o deplexión.

En estos transistores el canal está **preformado**, como se ve en las figuras 6-21 y 6-22 para los transistores de canal n o canal p respectivamente.

El canal se estrangula por deplexión, de forma análoga a como ocurría en los transistores FET. Los transistores conducen inicialmente y la corriente se controla con la puerta polarizando para aumentarla o para cortarla. Las figuras 6-23 y 6-24 son las características de estos transistores.

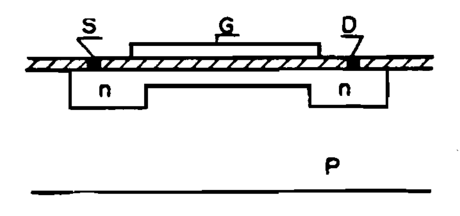


Fig. 6-21.

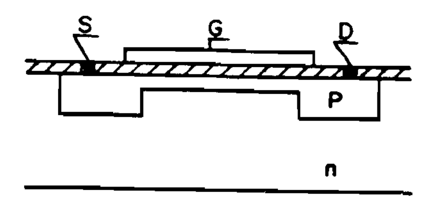


Fig. 6-22.

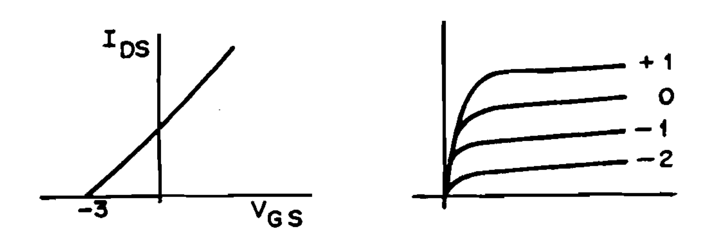


Fig.6-23.

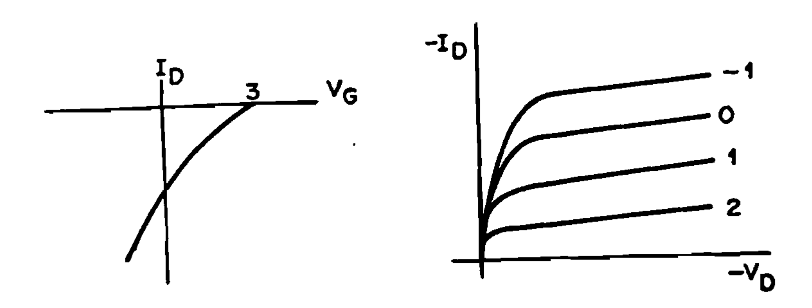


Fig.6-24.

6.13.- Transistores CMOS y Schottky.

La utilización más frecuente de los transistores es en conmutación, ya que tienen sobre los transistores dos ventajas muy importantes: La impedancia de entrada es muy alta y es independiente de la tensión de puerta. Especialmente adecuados son los MOSFET de enriquecimiento, que no conducen con tensión cero o inversa a la del drenaje en la puerta y conducen cuando la tensión de puerta es del mismo signo que el drenaje, cualquiera que sea su valor. Todavía se aprovecha mejor esta ventaja usando dos transistores de enriquecimiento, uno de canal n y otro de canal p **en serie**, con **puerta común**. El dispositivo es de integración muy fácil, como se ve en la figura 6-25, y se llama **CMOS**, MOS complementario.

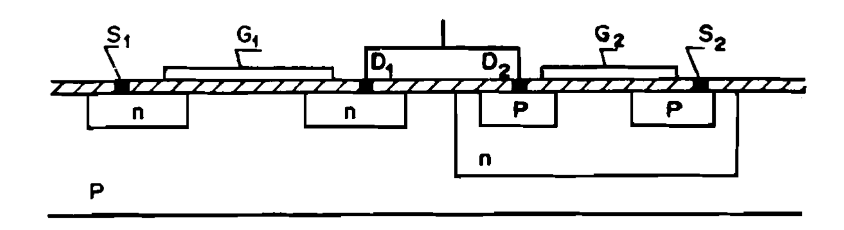


Fig. 6-25.

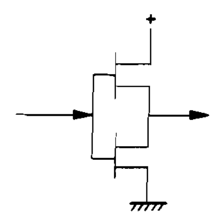


Fig. 6-26.

Si la polarización de la puerta es baja, cero, el transistor n no conduce y el transistor p conduce por estar fuertemente polarizado (fig. 6-26). La salida queda prácticamente conectada a la alimentación positiva, es alta. Si la polarización es alta (por ejemplo igual a la alimentación), el transistor n conduce y el p queda cortado. La salida queda conectada a tierra. Como veremos más adelante éste es el funcionamiento de un **inversor**.

Otra variante de FET es el **FET Schottky**, formado por una unión metal-semiconductor. Es un FET de empobrecimiento o deplexión en que la única unión es de tipo metal-semiconductor, o **diodo Schottky**. El diodo mismo tiene la ventaja de una menor tensión de saturación y el FET unipolar presenta muy poca capacidad y es favorable en algunas aplicaciones de alta frecuencia (fig. 6-27).

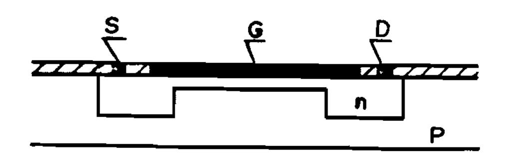


Fig. 6-27.

6.14.- Símbolos de los transistores FET y MOSFET.

La figura 6-28 muestra los símbolos de los diferentes tipos de MOSFET y FET. La flecha entrante denota un sustrato de tipo p y, saliente, de tipo n. En general se usan los símbolos de MOSFET simplificados. En ellos no se define la polaridad ni el tipo de MOSFET, pero es corriente que todos los transistores del sistema sean del mismo tipo, diciéndose entonces que se utiliza una "tecnología MOSFET de canal n", por ejemplo, para todo el sistema.

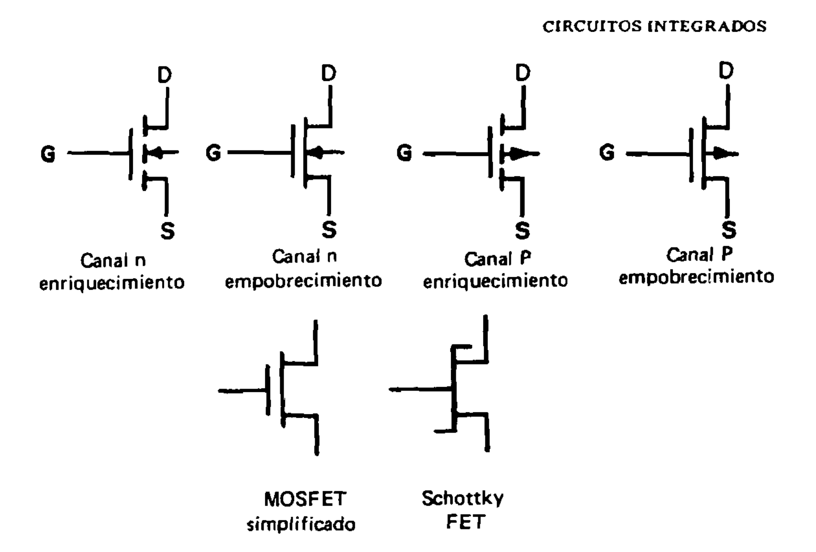


Fig. 6-28.

6.15.- Otros dispositivos semiconductores.

Existe una variedad de dispositivos basados en las técnicas de semiconductores dopados, algunos de los cuales han desaparecido por razones de mercado o por ser sustituidos por otros. Citaremos sólo algunos ejemplos:

a) Transistores con multielectrodos.

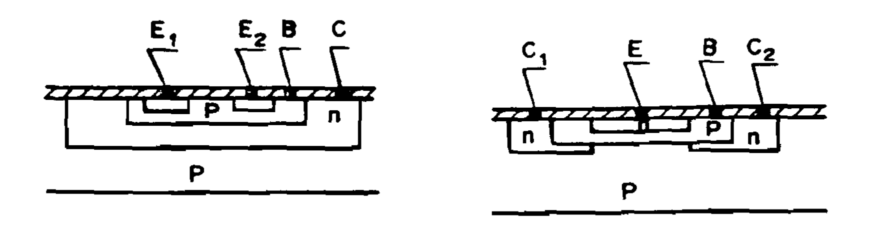


Fig. 6-29.

Usando técnicas de integración es fácil construir dispositivos como el transistor **multiemisor** o el **multicolector** de la figura 6-29, cuyo funcionamiento se comprenderá fácilmente cuando se estudien técnicas de conmutación.

De la misma forma se pueden construir MOSFET multipuerta. Se trata, como se aprecia en la figura 6-30 de dos transistores MOSFET integrados y conectados en serie. Se verá su aplicación cuando se estudien puertas lógicas.

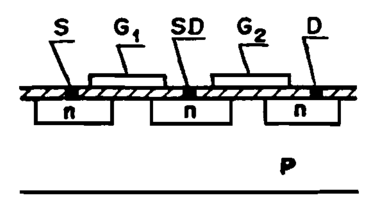


Fig. 6-30.

Todavía hay otras variantes, como MOSFET multifuente, multidrenaje, etc..., que pueden comprenderse fácilmente por analogía con los anteriores.

b) TRIAC, o conmutador controlado.

El SCR es un diodo, controlado en un sentido y que simplemente no conduce en el otro. Esta última propiedad es un inconveniente en muchas aplicaciones y puede evitarse con una disposición ingeniosa de las capas p y n de dopado, construyendo una nueva componente llamada TRIAC. La figura 6-31 representa la disposición de zonas en este dispositivo, así como la característica y el símbolo del mismo.

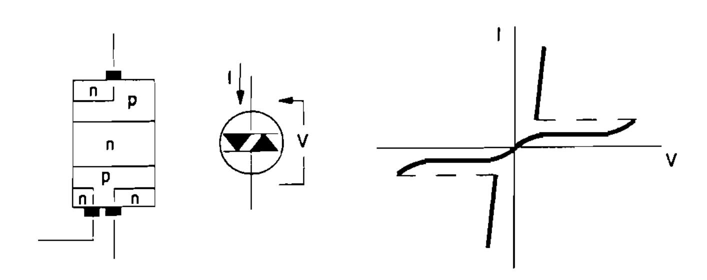


Fig. 6-31

Las flechas de la figura representan los trayectos pnpn en un sentido y en otro de la corriente convencional. La puerta única dispara con trigger de uno u otro sentido.

c) DIAC, doble diodo.

El DIAC es un doble diodo como se ve en la figura 6-32. Su

funcionamiento, que se aprecia en la característica es parecido al de un Zener y un diodo en ambas direcciones, marcando una tensión a la cual empieza a conducir bruscamente, como se aprecia en la figura. Tiene utilidad como dispositivo de disparo a un umbral determinado.

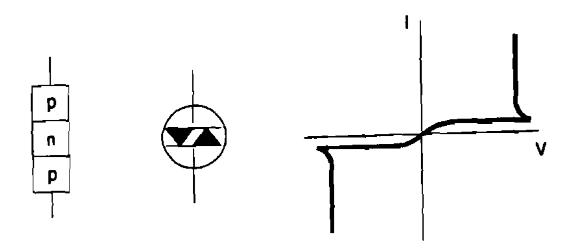


Fig. 6-32.

d) Transistor **bibase** o "unijunction" y bibase programable (**PUT**).

Es un dispositivo de canal, parecido a un FET, pero cuya puerta ataca al canal en unpunto a una altura determinada, que corresponde a una fracción determinada de la distribución de potencial a lo largo del canal, como se ve en la figura 6-33. Cuando la tensión en la puerta supera ese valor, conduce.

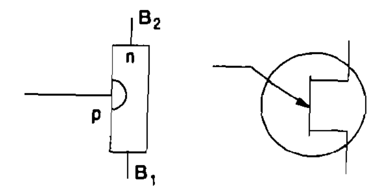


Fig. 6-33.

A veces se coloca una puerta que fuerza la distribución de potencial y entonces se llama "programable".

El uso de estos últimos dispositivos ha decrecido en los últimos años, sustituidos por técnicas digitales de mayor flexibilidad de uso, aunque no han desaparecido totalmente.

CAPITULO 7

TRANSDUCTORES

7.1.- Introducción.

La Electrónica o, más bien, los circuitos electrónicos manejan y transforman la información contenida en las señales eléctricas. Las variables que contienen la información están codificadas en estas señales eléctricas. Pero en los sistemas reales, las **entradas** del sistema rara vez son eléctricas, sino más bien variables físicas diversas, como **presión**, **temperatura**, etc... que es preciso transformar o **codificar** en forma eléctrica para que puedan ser manejadas por los circuitos. A su vez, las **salidas** del sistema no son útiles en forma eléctrica, sino que deben aparecer en forma de desplazamientos, o en forma óptica en instrumentos y presentaciones de diferentes tipos.

Los componentes que transforman las variables físicas en eléctricas se llaman **transductores**. Los que transforman las salidas eléctricas en desplazamientos y movimientos se llaman **actuadores** (normalmente basados en motores eléctricos) y los que dan la información en forma visual se llaman **presentaciones** (displays).

La denominación de transductores no está muy extendida en Electrónica; nosotros la usamos por su carácter general, pero, en la práctica, es más corriente usar las siguientes denominaciones:

Los **sensores** son los transductores que transforman o "miden" variables físicas como desplazamientos, velocidades, presión o temperatura y aceleraciones.

Los **micrófonos** son los transductores que transforman las señales acústicas en eléctricas.

Las **antenas** son dispositivos que transforman señales electromagnéticas en eléctricas y viceversa. Por su importancia en comunicaciones les dedicaremos un capítulo especial.

Los sistemas **Selsyn** miden señales angulares.

Los **giróscopos** miden posiciones y velocidades angulares.

Existe alguna confusión en las denominaciones de los transductores, que no tendremos en cuenta en este capítulo. Estudiaremos algunos tipos de transductores subrayando, en cada uno de ellos, las variables de entrada y salida y los circuitos auxiliares de medida.

7.2. Sensores de posición.

Generalmente se usan sensores de posición para medir deformaciones de piezas o estructuras o posición de una parte móvil de un sistema. Los desplazamientos moderadamente grandes se miden por medio de potenciómetros o resistencias variables en las que la posición del cursor se transforma en variación de resistencia. Los desplazamientos mayores se transforman en pequeños por medio de los mecanismos adecuados.

Los pequeños desplazamientos se transforman en eléctricos mediante **"captores"** inductivos o capacitivos como los representados en la figura 7-1.

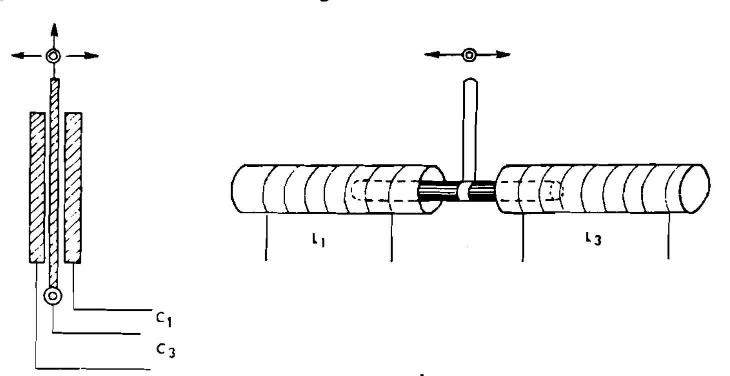


Fig. 7-1

En la figura de la izquierda, el movimiento de la pieza móvil desequilibra la igualdad de dos condensadores, aumentando la capacidad del condensador en el que se acercan sus dos electrodos y disminuyendo la capacidad del otro. Con esta medida **diferencial** se compensan algunos errores debidos a perturbaciones simétricas, ya lo que varía al cambiar la posición es la diferencia de capacidades. En el captor representado a la derecha de la figura, son dos autoinducciones las que varían al moverse la pieza. Las dos autoinducciones son iguales cuando la posición es central y su diferencia es cero. Esta diferencia es la verdadera **variable de salida**, con lo que el captor puede también considerarse diferencial.

Existen muchos otros transductores que primero traducen la variable a medir en un desplazamiento para luego medir éste. Así son los dinamómetros que miden la extensión de un muelle o los

captores de presión o carga de los materiales.

7.3.- Acelerómetros.

Los acelerómetros son un tipo especial de **dinamómetro**, que miden el desplazamiento de una masa de inercia suspendida de un muelle (fig. 7-2). La aceleración a que está sometido el acelerómetro produce una fuerza de inercia sobre la masa **M** que es igual al producto de la masa por la aceleración en sentido contrario a la aceleración.

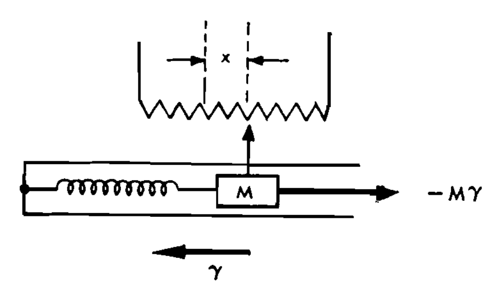


Fig. 7-2

La proyección de esa fuerza sobre la dirección de desplazamiento del émbolo produce un alargamiento x del muelle proporcional a Mγ, siendo γ la proyección de la aceleración sobre el eje del acelerómetro. El desplazamiento se transforma en una variable eléctrica arrastrando un potenciómetro, como se indica en la figura. El cambio de resistencia, proporcional al desplazamiento, lo es también a la aceleración. El acelerómetro, como elemento mecánico, responde a los transitorios a través de una ecuación de segundo orden que depende de la fricción y de la constante del muelle. El comportamiento dinámico del acelerómetro se estudiará más adelante.

7.4.- Captores de presión.

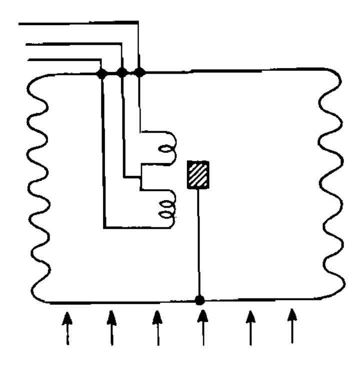


Fig. 7-3

La figura 7-3 representa un sensor o captor de presión del tipo de los utilizados para medir presión hidráulica o nivel de líquido. Se trata de un dinamómetro en que una membrana elástica convierte la presión en desplazamiento de un "fiel" que desequilibra la igualdad entre dos autoinducciones, moviendo un núcleo magnético. El conjunto se encapsula en un "bote" hermético que se sumerge en el líquido cuyo nivel se quiere medir. La diferencia de autoinducciones de las dos bobinas se mide con un **puente de medida** como el que describiremos posteriormente.

7.5.- Sensores de temperatura.

Es fácil medir la temperatura a través de su influencia en algunas variables eléctricas.

Las resistencias metálicas son un medidor natural de temperatura, ya que su valor aumenta casi linealmente con la temperatura, a través de un "PTC", coeficiente de variación con la temperatura positivo, lo que las convierte directamente en **termistores** o medidores de temperatura.

Los materiales semiconductores varían su resistividad con la temperatura disminuyéndola, es decir, poseen un "NTC" o coeficiente de variación de la resistencia con la temperatura negativo, en general de valor considerable, aunque no lineal. Conociendo el valor de esta variación, una resistencia de material semiconductor es también un termistor NTC.

También se puede utilizar como medida de la temperatura la variación de la tensión de un diodo a corriente constante, o la variación de la tensión de base de un transistor a corriente constante, que es de unos -2 mV / ºC.

7.6.- Otros sensores.

También se pueden medir otras variables físicas por su efecto sobre variables eléctricas. Citaremos a continuación algunos ejemplos:

La **luz** se puede medir con **fotoresistencias** y **fotodiodos**. En ambos casos se trata de semiconductores en los que la incidencia luminosa aporta energía a las cargas libres aumentando la conductividad.

La **humedad** cambia la resistencia de los materiales aglomerados porque aumenta la concentración de iones salinos entre partículas o en suspensión. Los transductores de humedad suelen ser circuitos de electrodos impresos sobre una placa de material poroso. Al aumentar la humedad, el material poroso retiene agua y aparece una conductividad entre los electrodos.

El **pH** de un líquido influye sobre la tensión de contacto de los electrodos, que es lo que se puede medir con amplificadores de gran impedancia de entrada (FET, por ejemplo).

7.7.- Micrófonos.

El micrófono es un transductor que transforma las ondas sonoras (de desplazamiento - presión del aire) en señales eléctricas. En cierto modo se trata de un dinamómetro o medidor de presión que mide variaciones muy pequeñas de la presión. Por eso los micrófonos son variaciones sobre los transductores anteriores.

El **micrófono de condensador** (figura 7-4) es un medidor de presión capacitivo. Las variaciones de presión del aire se transforman en pequeños desplazamientos de una membrana y estos movimientos se transforman en variaciones de capacidad entre la membrana y otro electrodo metálico.

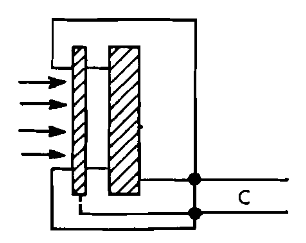


Fig. 7-4.

El **micrófono de condensador** tiene una respuesta de frecuencia y fidelidad muy satisfactorias, por lo que se usa en medidas acústicas de precisión. Sin embargo, la señal obtenida es muy pequeña y necesita circuitos auxiliares complicados.

El **micrófono de carbón** utiliza un aglomerado de polvo de carbón cuya resistencia eléctrica cambia con la presión. No es muy preciso pero produce señales grandes y es muy barato. Se utiliza en telefonía y en otras aplicaciones de poca calidad.

El **micrófono dinámico** es un captor de presión inductivo como los representados en la figura 7-5. En esta figura se representa un micrófono de bobina móvil (el más usado) y otro de imán móvil. En ambos casos el captor tiene un imán que induce fuerzas electromotrices en una bobina al desplazarse ésta respecto a él.

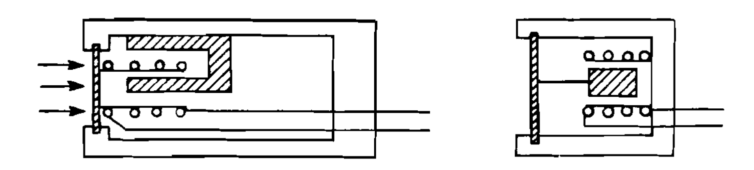


Fig. 7-5

7.8.- Variable de salida y medida.

En algunos casos la variable de salida es una tensión eléctrica, como en las células fotoeléctricas o en el micrófono

dinámico, pero en muchos otros la salida del sensor es una variación de resistencia o de capacidad o autoinducción. En estos casos se necesita un dispositivo de medida o **puente de medida**.

Cuando la salida es la posición de un potenciómetro, se transforma fácilmente en tensión, aplicando una tensión constante al potenciómetro que se convierte en un divisor de tensión. La figura 7-6 es un ejemplo de esta conversión. Se aplica una tensión E entre los extremos del potenciómetro de resistencia R, bobinado sobre un sector angular β. Al girar un ángulo α, la tensión de salida será la que corresponde a la fracción $\alpha R/\beta$, o sea $\alpha E / \beta$.

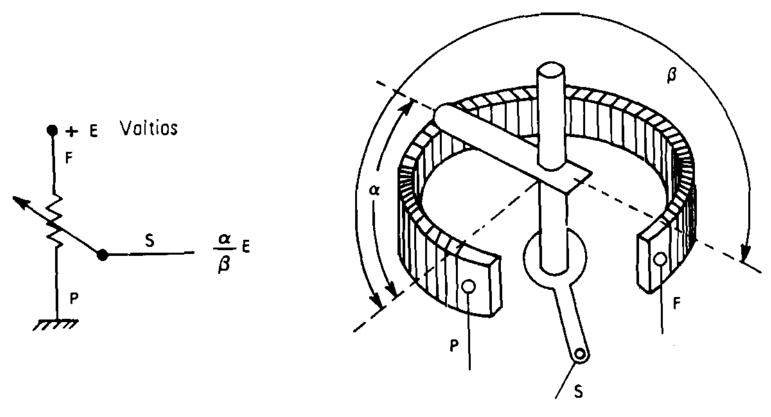


Fig. 7-6.

Si la salida es una resistencia variable, como en el caso de los sensores de temperatura, por ejemplo, se puede medir con un puente de Wheatstone como el de la figura 7-7, en la que R_1 es la resistencia sensible a la temperatura, R_3 se ajusta para equilibrar el puente al valor que elijamos como origen o a un valor central (para medir pequeños desplazamientos a partir de ese punto) y las otras dos resistencias son arbitrarias.

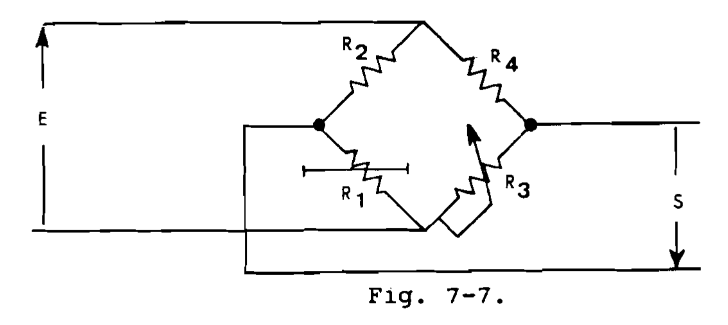


Fig. 7-7.

En el valor de origen, R_3 se ajusta para que:

$$\frac{R_1}{R_2}=\frac{R_3}{R_4}$$

Cuando la temperatura ha cambiado, es decir, R_1 ha cambiado, la salida del puente es la diferencia entre dos divisores de tensión. Esa salida será:

$$S=E[\frac{R_1}{(R_1+R_2)}+\frac{R_3}{(R_3+R_4)}]$$

La variación puede considerarse lineal respecto a R_1 si se desprecia su variación en el denominador. En todo caso puede tenerse en cuenta la dependencia a través de la fórmula para medir la temperatura.

Hay que hacer notar que si el transductor es una capacidad o una autoinducción puede hacerse el mismo razonamiento con las impedancias de la figura 7-8 que con las resistencias del caso anterior.

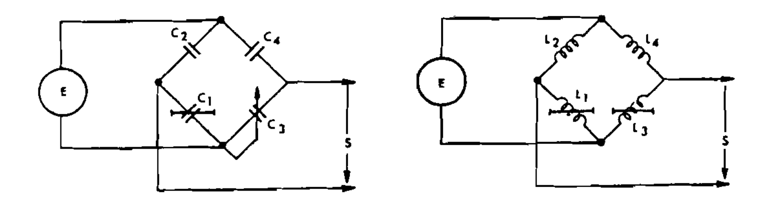


Fig. 7-8.

A continuación calcularemos la **sensibilidad** del puente de medida, es decir la variación relativa de la tensión de salida respecto a la variación relativa de impedancia del transductor.

Derivando la expresión de E en función de R_1 obtenemos:

$$\frac{dS}{dR_1}=E\frac{R_2}{(R_1+R_2)^2}$$

multiplicando ambos miembros de la igualdad por R_1 y dividiendo por E, se obtiene :

$$\frac{\frac{dS}{E}}{\frac{dR_1}{R_1}}=\frac{R_1R_2}{(R_1+R_2)^2}$$

Si mantenemos R_1 + R_2 constante (es decir, carga constante

para la alimentación), el valor óptimo del producto con suma constante se obtiene para $R_1 = R_2$ y entonces el valor es:

$$\frac{\frac{dS}{E}}{\frac{dR_1}{R_1}} = \frac{R^2}{4R^2} = \frac{1}{4}$$

que nos dice que la variación relativa de tensión es 1/4 de la variación relativa del transductor.

La ventaja más importante del puente de Wheatstone es que las perturbaciones que afecten igualmente a R_1 y a R_3 se compensan entre sí. Si, por ejemplo, se quiere medir luz incidente con un transductor sensible a luz y temperatura, pueden ponerse dos transductores iguales en R_1 y R_3 y someter ambos a la misma temperatura pero hacer incidir la luz sólo sobre R_1. Las variaciones de temperatura se compensarán y las de luz incidente no.

7.9.- Giróscopos.

Un giróscopo es una masa o volante que se hace girar a alta velocidad (generalmente a más de 20.000 r.p.m.) y que se suspende en un chasis con uno o dos grados de libertad a través de una suspensión adecuada. Por ejemplo, si, como se ve en la figura 7-9, el eje de giro es horizontal, perpendicular al papel, el "cuadro" de sujeción del eje se sujeta mediante una suspensión giratoria de eje horizontal a otro cuadro que puede girar sobre un eje vertical. Es decir, en este caso las dos libertades de movimiento se obtienen sobre los dos ejes perpendiculares al eje del volante. Si se restringe la libertad de movimiento sobre uno de esos ejes se obtiene el giróscopo de un grado de libertad.

Del estudio mecánico del giróscopo se deduce la **ecuación del giróscopo**, que relaciona la velocidad angular sobre un eje (perpendicular al eje del volante) con el momento aplicado sobre el tercer eje (perpendicular a los otros dos ejes). La ecuación el giróscopo es :

$$\overline{M} = [\overline{\Omega} \times \overline{\omega}]$$

es decir, el par aplicado en uno de los ejes perpendiculares al eje del volante es igual al producto vectorial del momento angular del volante por la velocidad angular sobre el tercer eje.

La demostración de esta ecuación, análoga a la ley de Newton, puede encontrarse en la literatura especializada. Aquí sólo comentaremos sus peculiaridades:

En primer lugar, si se aplica un par M al giróscopo, éste reacciona con una **velocidad** ω (y no una aceleración) **perpendicular** (y no en la misma dirección) al momento aplicado. Además, la reacción es inversamente proporcional al momento angular del volante Ω, que puede ser muy grande si, como ocurre en la práctica, el volante gira a muchas revoluciones. El giróscopo tiene una inercia aparente mucho mayor que la que corresponde a su masa, y de ahí surgen algunas de sus aplicaciones. Por otra parte, si se le da una **velocidad angular** (equivalente a una aceleración en la ley de Newton), aparece como reacción un **momento perpendicular**, que puede utilizarse también en otras aplicaciones.

La velocidad angular de respuesta a un par aplicado se llama **precesión** y es lo que se observa al hacer girar una peonza inclinada sobre un suelo plano. El par de reacción del suelo hace desplazarse al peón lateralmente en un movimiento en círculos que todos hemos observado.

Concretemos ahora algunas aplicaciones aeronáuticas del giróscopo de **dos** grados de libertad:

La figura 7-9 representa un **girodireccional**. El giróscopo, suspendido con dos grados de libertad, mantiene su orientación cuando el avión al cual está sujeto el eje vertical gira en el sentido del azimut o **rumbo**. Debido a su gran momento de inercia, los pequeños pares de fricción de arrastre no lo mueven apreciablemente. El cuadro exterior indica respecto al avión el ángulo girado, que puede leerse directamente o transformarse en señal eléctrica moviendo el cursor de un potenciómetro.

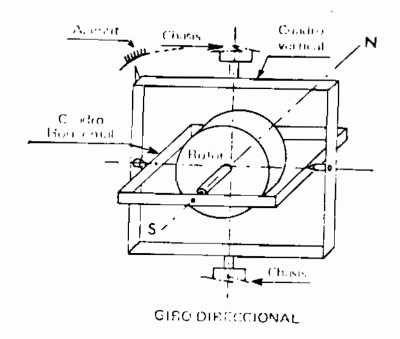


Fig. 7-9

La figura 7-10 es un **horizonte artificial**, basado en un efecto parecido. El eje del volante conserva su orientación absoluta a pesar de los movimientos de **picado** del avión en el plano vertical. El indicador o el transductor angular correspondiente indican la posición de **asiento** del avión. También puede observarse directamente el cuadro horizontal, que conserva su posición cuando el avión pica o se encabrita, dando la sensación de una visión de horizonte, de donde recibe su nombre el dispositivo.

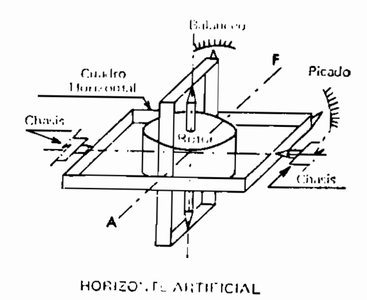


Fig. 7-10

La figura 7-11 representa un **medidor de velocidad angular** con un giróscopo de un grado de libertad. En este caso se trata de medir la velocidad de **guiñada** (movimiento en el plano horizontal, a derecha e izquierda). La velocidad en ese plano, con eje vertical, produce un par en el tercer eje, perpendicular al papel, que se mide con un muelle cuya deformación será proporcional al par y, por consiguiente, a la velocidad de guiñada. Esta desviación puede medirse en forma eléctrica con un potenciómetro.

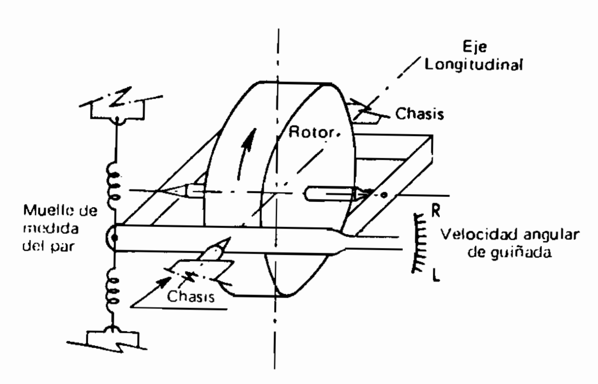


Fig. 7-11

Una aplicación que ha caído en desuso es la provocación de una precesión aplicando un par proporcional a una variable función del tiempo. La velocidad angular de precesión es proporcional a la variable de entrada y la posición angular (integral de la velocidad) será proporcional a la integral de la variable de entrada. Se obtiene así un **integrador mecánico**. Este tipo de integrador fue usado en el sistema de guiado de los célebres misiles V-2 de la segunda guerra mundial. Actualmente, la integración se hace con circuitos electrónicos.

7.10.- Sistemas Selsyn.

El sistema Selsyn es un transductor **repetidor** angular clásico basado en la teoría de máquinas eléctricas, con la que

hay una larga experiencia. Los sistemas Selsyn son de dos clases que describiremos a continuación:

El sistema **Selsyn pasivo** está constituido por una bobina móvil, alimentada por una fuente de corriente alterna, que se orienta mecánicamente según el ángulo a medir o repetir. La figura 7-12 reproduce un **generador selsyn**. La bobina genera un campo magnético variable que induce fuerzas electromotrices en las bobinas del **estátor** o parte fija del dispositivo. Las bobinas fijas (tres en la figura, pero pueden ser dos o cuatro) actúan como secundarios de tres transformadores. El acoplamiento y, a través de él, la fuerza electromotriz o tensión inducida en cada una de las tres bobinas es proporcional a los cosenos directores, o cosenos de los ángulos formados por la bobina móvil y las fijas. Por todo ello se llama a este Selsyn transductor angular **transmisor Selsyn**.

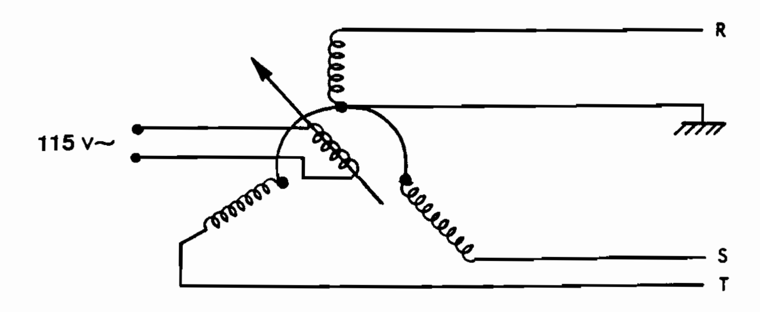


Fig. 7-12.

La información del ángulo de situación de la bobina fija está contenida en los tres valores de tensión de salida de las tres bobinas. Para valorarla en el sistema pasivo se usa otro **selsyn receptor** exactamente igual al anterior pero usado como receptor, como se ve en la figura 7-13.

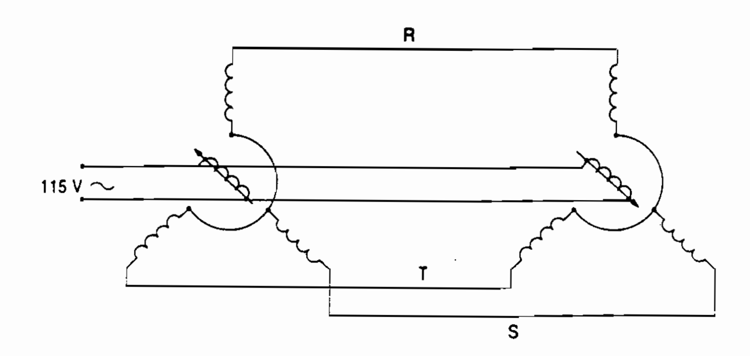


Fig. 7-13.

Las tres tensiones del transmisor suministradas a las tres bobinas análogas en el receptor reproducen la dirección del campo

magnético alternativo de la bobina móvil del transmisor. La bobina móvil del receptor se alimenta con la misma fuente que la del transmisor, con lo que también produce un campo magnético propio igual al de la bobina móvil del transmisor. Si ahora se la deja en libertad de girar, tiende a orientarse en dirección opuesta a la del campo magnético de las bobinas fijas, es decir, reproduce la dirección angular de la bobina móvil del transmisor con 180º de desfase. El selsyn receptor se convierte así en un **repetidor angular** de la posición del transmisor, dando una indicación a distancia de la misma.

El inconveniente del repetidor pasivo es que el par motor del receptor no puede ser muy grande teniendo en cuenta que la bobina móvil del transmisor debe ser ligera para no perturbar la medida que está haciendo. Para trasmitir la posición angular a distancia con mayor potencia se usa el **selsyn activo** que es un servomecanismo con realimentación (figura 7-14).

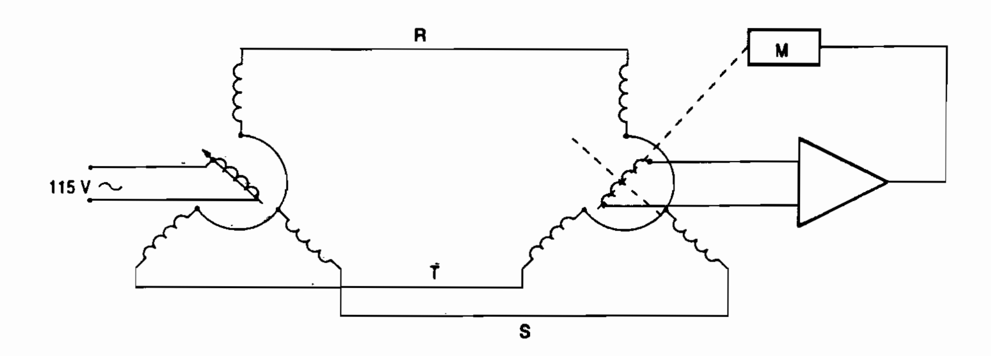


Fig. 7-14.

El transmisor del selsyn activo funciona exactamente igual que el pasivo, es decir, la bobina móvil se alimenta con una fuente de c.a. y genera tensiones inducidas en las bobinas del estátor que son proporcionales a los "cosenos directores" que la bobina móvil forma con ellas. Igual que en el selsyn pasivo, en el receptor se reproduce el campo magnético de la bobina móvil del transmisor, producido por las bobinas fijas. Pero en el selsyn activo no se alimenta la bobina del receptor, sino que se usa como instrumento de medida angular : La fuerza electromotriz inducida en la bobina móvil del receptor será máxima cuando ésta esté orientada en la dirección del campo, o sea en la misma orientación de la bobina transmisora, y será cero cuando se oriente en sentido perpendicular al campo. Precisamente esta posición, la de fuerza electromotriz inducida cero, es la que se usa como correcta en el receptor, al que se obliga a seguir siempre en cuadratura al transmisor (la indicación del ángulo trasmitido será la perpendicular a la bobina). La desviación de esa posición es la **indicación de error** del servosistema. Esta medida de error se amplifica y se lleva a mover un motor que obligue a la bobina móvil a girar en el sentido de anular el error. Cuando la bobina móvil llegue a su posición correcta

(perpendicular al transmisor), el error se anula y el motor se para. Con este sistema activo se puede arrastrar cualquier dispositivo mecánico solidario con la bobina móvil con la potencia del motor (y no con la del campo del selsyn). La estabilidad de este sistema de regulación de posición **activo** se estudiará en un capítulo posterior.

CAPITULO 8

AMPLIFICACION LINEAL

8.1.- Concepto de amplificador.

Al estudiar los cuadripolos hemos definido la **amplificación** como relación entre la señal de salida y la de entrada. Ahora vamos a detallar algo más este concepto, teniendo en cuenta también la **potencia de señal**.

No existe dificultad alguna en la definición de potencia de señal cuando ésta viene en forma de una tensión entre los extremos de una impedancia. La **potencia de entrada** suele definirse como la potencia eléctrica que "aparece" sobre la impedancia de entrada al llegar la señal, procedente del exterior o del circuito anterior. En el caso más corriente de que la impedancia de entrada sea resistiva de valor R_E, la potencia de entrada cuando aparece una señal de valor eficaz E_E será E_E^2/R_E.

Sin embargo, no existe unanimidad sobre la definición de **potencia de salida** de un circuito. Una definición lógica sería la potencia que el circuito suministra sobre la carga. Esta definición tiene el inconveniente de que es necesario definir la carga y la potencia de salida deja de depender sólo del amplificador para pasar a ser un parámetro del amplificador con su carga. Este inconveniente puede resolverse definiendo una **carga nominal** sobre la que se aplica la señal de salida, o incluyendo la carga en el amplificador. Otro criterio puede ser tener en cuenta la resistencia de salida del amplificador y tomarla como carga. Como se recordará, la máxima potencia que se puede extraer de un generador ocurre cuando la carga exterior es igual a la resistencia de salida. Sin embargo, la tensión de salida en ese caso y, con ella, la amplificación, se reduce a la mitad. En todo caso, la relación entre la potencia de salida y la potencia de entrada se llama **amplificación de potencia** y su

valor será:

$$A_W=\frac{\frac{E_S^2}{R_L}}{\frac{E_E^2}{R_E}}=A_L^2\frac{R_E}{R_L}$$

donde A_L es la amplificación con carga nominal R_L, R_E es la resistencia de entrada y R_L la resistencia de carga nominal. Si, en lugar de la carga nominal tomamos la resistencia de salida, la fórmula se convierte en :

$$A_W=A_L^2\frac{R_E}{R_S}=\frac{A_0^2}{4}\frac{R_E}{R_S}$$

teniendo en cuenta que la amplificación con carga igual a la resistencia de salida es la mitad de la amplificación sin carga.

A veces se define la amplificación sin tener en cuenta el factor 1/4, como si la resistencia de salida fuese la nominal y el amplificador fuese insensible a la carga, lo que es cierto para amplificadores fuertemente realimentados. En todo caso, la primera fórmula es siempre válida, sea cualquiera la carga asociada a la definición.

Un **amplificador** se caracteriza porque $A_W > 1$, es decir, porque la amplificación de potencia es mayor que 1. En general, es también la amplificación de tensión A_0 mayor que 1, aunque esto no es necesario y no se cumple en los "adaptadores de impedancia", que veremos más adelante. También existen dispositivos en los que $A_0 > 1$ y $A_W < 1$ y **no** son amplificadores (por ejemplo, transformadores y circuitos resonantes).

Normalmente, el parámetro más usado en un amplificador será A_0, dando por supuesto que $R_E >> R_S$ y que, si A_0 es alto, también lo serán A_L y A_W.

8.2.- El transistor como amplificador.

El transistor es un dispositivo que controla la corriente de emisor- colector aportando una pequeña potencia entre el emisor y la base. La **salida** será pues el colector o el emisor y la **entrada** la base o el emisor. Según los electrodos que se utilizan como entrada y salida, existen tres tipos de amplificadores con un transistor:

a) **Emisor a masa**. La entrada es la base y la salida , el colector. El emisor se utiliza como electrodo de referencia, aunque no es necesario conectarlo directamente a masa, como

veremos. Es la conexión más usada por su alta amplificación y baja impedancia de entrada (figura 8-1) y a ella nos referiremos siempre cuando no se indique lo contrario.

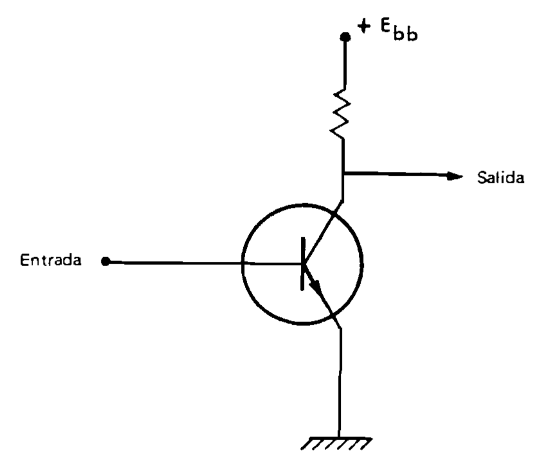


Fig. 8-1.

b) **Colector a masa**. La entrada es la base y la salida, el emisor. El colector se conecta a la alimentación, lo que es equivalente a masa para pequeñas señales. Se usa como adaptador de impedancias por su amplificación próxima a la unidad, alta impedancia de entrada y baja impedancia de salida. Se llama **seguidor de emisor**, porque el emisor "sigue" a la base con amplificación próxima a uno (figura 8-2).

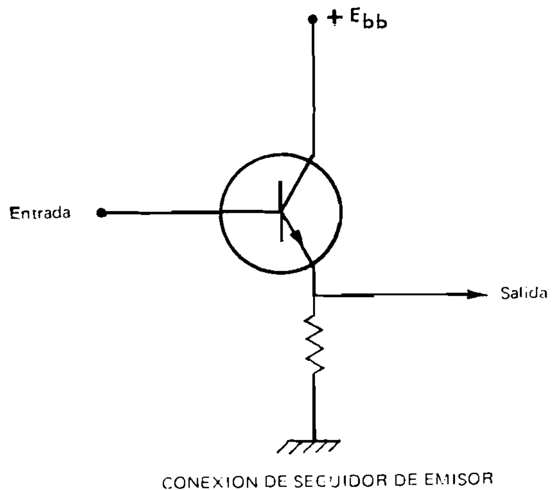


Fig. 8-2

c) **Base a masa**. La entrada es el emisor y la salida el colector. Sus características de amplificación son parecidas a

a), con el inconveniente de una mucho menor impedancia de entrada, que presenta algunas ventajas, por su bajo nivel de ruido, en etapas de entrada de sistemas muy sensibles al ruido, como el radar (figura 8-3).

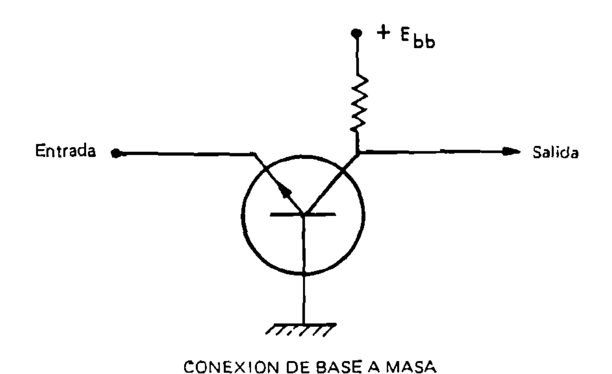


Fig. 8-3

En lo sucesivo nos referiremos siempre a la conexión de emisor a masa, cuando no se especifique lo contrario. Sólo estudiaremos circuitos con transistores npn, aunque los resultados se generalizan muy fácilmente a los pnp si se desea.

8.3.- Polarización y señales.

Para que un transistor funcione como tal, es decir, controlando la corriente de colector como hemos visto en el capítulo 5, es necesario que esté en **zona de trabajo**. Se llama zona de trabajo al conjunto de valores de las variables en que las ecuaciones son válidas. Por ejemplo, si la tensión base-emisor en un transistor npn es negativa, el transistor no conduce porque el diodo base emisor recibe una polarización inversa y la tensión de base misma no controla nada, porque su valor influye muy poco o nada en la corriente de colector.

Hay muchos dispositivos técnicos en que ocurre lo mismo. Sirva como ejemplo el acelerador de un automóvil: Si se pisa a fondo, la mariposa está abierta al máximo, independientemente de que el pedal esté más o menos a fondo. El conductor experto sabe que hay una zona de control, que es la zona en que la mariposa recorre el margen de estrangulamiento del flujo y, fuera de esa zona, deja de controlar. Lo mismo ocurre con el embrague.

La operación de dar unos valores iniciales, que pongan al transistor en zona de trabajo, se llama **polarizar**. A partir de ese momento se aplican las señales en forma de pequeñas desviaciones desde ese conjunto de valores iniciales. Al conjunto de valores iniciales se le llama **punto de trabajo**, que, naturalmente, estará dentro de la **zona de trabajo**. A partir del punto de trabajo, la información se da en forma de **desviaciones** desde ese punto. Estas desviaciones no deben "sacar" a las variables de la zona de trabajo, para que las ecuaciones sigan

siendo válidas.

Puede ocurrir que se desee que el transistor se salga de la zona de trabajo, o bien en algunos casos excepcionales (por ejemplo, para limitar las señales), lo que estudiaremos en otro capítulo como **amplificación no lineal**, o más frecuentemente cuando utilicemos el transistor en **conmutación** o en **técnicas digitales**, que también estudiaremos más adelante. Pero, generalmente, cuando hablamos de amplificadores nos referimos a **amplificadores lineales**, que funcionan a partir del punto de trabajo y con **señales pequeñas**, que se desvían poco del punto de trabajo y, además, permiten linealizar las ecuaciones. En este capítulo estudiaremos sólo este tipo de amplificadores de pequeñas señales, o lineales.

En adelante, denominaremos con mayúsculas los valores del punto de trabajo, a los que también se les llama de **reposo** o de **corriente continua**, en el sentido de que no varían hasta la aparición de las señales. A las señales las denominaremos con minúsculas.

Resumiendo, podemos decir que en cualquier problema de amplificación lineal debemos considerar:

a) Un **punto de trabajo** constituido por un juego de valores de las variables que satisfacen las ecuaciones de los componentes y que, además, admiten pequeñas variaciones sin dejar de satisfacerlas. Al estudio de este punto de trabajo lo llamaremos de **estado de reposo** o de **corriente continua (c.c.)** o también **no lineal**.

b) Unas **señales** que se superponen a los valores anteriores de forma que las variables permanezcan en zona de trabajo. Además, estas desviaciones deben ser lo suficientemente pequeñas para que las ecuaciones se puedan **linealizar**. Al análisis de señales (por ejemplo, amplificación, impedancias de entrada y salida) le llamaremos también **lineal** o de **señales** o de **corriente alterna (c.a.)**, ya que las señales suelen ser de c.a.

8.4.- Amplificación con un transistor.

El estudio de un amplificador lineal (pequeñas señales) con un transistor puede dividirse en tres partes :

1.- Análisis del circuito para calcular el punto de trabajo del transistor y las variables de polarización. Se analizará un circuito típico.

2.- Síntesis o proyecto del circuito.. Se calcularán los valores de las componentes para que el transistor funcione en el punto de trabajo deseado.

3.- Parámetros técnicos del amplificador : amplificación, impedancias de entrada y salida y realimentación de corriente. Equivale al estudio lineal o de pequeñas señales.

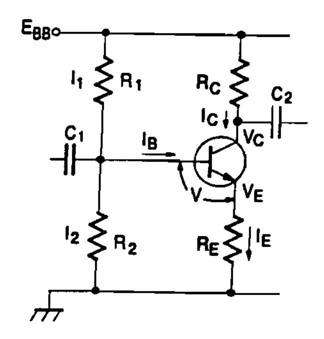


Fig. 8-4.

El circuito típico del amplificador es el de la figura 8-4. La resistencia R_C es la de **carga** y tiene la misión de convertir las variaciones de intensidad de colector en variaciones de tensión. La resistencia R_E es estabilizadora, como veremos después. El divisor de tensión formado por R_1 y R_2 tiene por objeto polarizar al transistor en su punto correcto. Los condensadores C_1 y C_2 deben dejar pasar las señales y separar la polarización. Las resistencias de colector y emisor pueden ser sustituidas por impedancias para obtener determinados resultados con las señales. En ese caso, los valores a tomar en el estudio no lineal serán los de la componente resistiva para cc. de dichas impedancias.

8.5.- Análisis no lineal. Punto de trabajo.

Llamemos I_E, I_C, I_B, V a las variables del transistor y V_C, V_E, E, I_1 e I_2 a las restantes del circuito.

Las ecuaciones del transistor son :

$I_E = K\ e^{39V}$ (Suponemos q/KT = 39 $volt^{-1}$ a temperatura ambiente)

$$I_C = \alpha\ I_E + I_{co}$$
$$I_B = (1 - \alpha)\ I_E - I_{co}$$

El circuito aporta las ecuaciones:

$E = V + V_E$ (Ecuación de "entrada").

$V_C = E_{BB} - I_C\ R_C$ (Ecuación de "carga").

$$V_E = R_E\ I_E$$
$$E_{BB} - E = R_1\ I_1$$
$$E = R_2\ I_2$$
$$I_1 = I_2 + I_B$$

Los parámetros que intervienen en las ecuaciones son los del transistor y la alimentación E_{BB}.

El problema está determinado con nueve ecuaciones (una de ellas no lineal) y nueve variables. Con su solución se determina el punto de trabajo del transistor y la solución es válida si éste está en la zona de validez de las características.

La solución puede hacerse de forma gráfica, como ya vimos anteriormente. Sin embargo, rara vez se plantea el problema de esta forma, ya que lógicamente, se **proyecta** el circuito para un punto de trabajo preelegido y no a la inversa.

La ecuación de entrada puede escribirse $E = V + R_E I_E$, donde V es función de I_E a través de la primera ecuación del transistor, donde la temperatura interviene a través de K y del exponente. Si hacemos el primer término (V) despreciable frente al segundo ($V << V_E$) resulta I_E independiente de V, igual a E/R_E, independientemente de la temperatura. Vemos así el papel estabilizador de R_E, que es el que introduce este término.

8.6.- Proyecto de circuito amplificador.

Normalmente se eligen las resistencias y componentes del circuito para cumplir las especificaciones deseadas, entre las que se cuenta que el transistor funcione en un punto de trabajo elegido.

El punto de trabajo se elige a la vista de las características del transistor (zona lineal, alejándose de saturación y curvaturas), lo que, junto con las ecuaciones del transistor, da el conjunto de valores I_E, I_C, I_B, V. A veces se eligen los valores recomendados por el fabricante, sin necesidad de consultar las características. Previamente se ha comprobado que la tensión máxima que soporta el transistor es superior al valor de la alimentación E_{BB}, que es un parámetro dado por el sistema.

V_C tiene un margen de elección muy grande y suele fijarse en las proximidades de $E_{BB}/2$, para alejarse de la saturación y de la alimentación, es decir, para tener el mayor margen de variación posible de la tensión de salida.

La ecuación de carga nos permite ahora calcular

$$R_C = (E_{BB} - V_C) / I_C.$$

R_E es una resistencia estabilizadora, como veremos después, perjudicial para la amplificación, por lo que debe ser pequeña respecto a R_C, por ejemplo del orden de 0,1 R_C. Más adelante se verá que esta estimación limita la amplificación a bajas frecuencias a 10, aunque es posible paliar esta limitación a altas frecuencias como ya veremos.

El divisor de tensión de entrada tiene todavía un grado de libertad. Si se eligen los valores de las resistencias altos se tiene un divisor con alta impedancia interna (recordar Thevenin) y sensible a las variaciones de I_B. Si se eligen los valores

bajos, se baja la impedancia de entrada para señales, con la que está conectado en paralelo. Un compromiso es elegir $R_2 = R_C$.

Ahora es fácil calcular $I_2 = E/ R_2$. $I_1 = I_2 + I_B$ y

$$R_1 = (E_{BB} - E)/ I_1.$$

8.7.- Parámetros técnicos del amplificador con un transistor.

La figura 8-5 es el circuito equivalente para señales de un amplificador con un transistor.

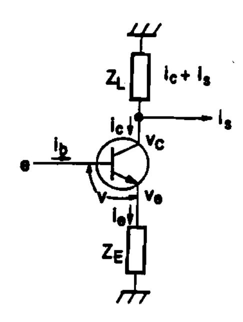


Fig. 8-5.

En relación con la figura 8-4 pueden observarse las siguientes diferencias :

Se ha prescindido del divisor de tensión de polarización de la base. Simplemente, la impedancia de entrada se calcula a partir de la base y el circuito de polarización se incluye posteriormente (el de la figura 8-4 u otro, en su caso).

Se han generalizado las resistencias de emisor y carga a impedancias complejas. Se entiende que se restringe el campo de aplicación de las señales a señales sinusoidales.

Se incluye una intensidad de salida que circula a través del acoplamiento hacia la carga exterior.

La alimentación E_{BB} se sustituye por masa (0), ya que las constantes tienen variación o señal cero.

Las variables se expresan con minúsculas y representan pequeñas variaciones de sus mayúsculas correspondientes, o bien directamente o a través de la transformación $j\omega$.

Las ecuaciones son ahora:

Ecuaciones del transistor:

$$i_e = 39\ I_E\ v$$

$i_c = \alpha\ i_e$
$i_b = (1 - \alpha)\ i_e$

Ecuaciones del circuito:

$v_c = -\ Z_C\ (i_c + i_s)$ (Ecuación de carga)
$e = v + Z_E\ i_e$ (Ecuación de entrada)
$v_e = Z_E\ i_e$

Tenemos seis ecuaciones con las variables de entrada del cuadripolo, e, i_b, de salida del cuadripolo, v_c , i_s e internas del cuadripolo i_e , v, i_c y v_e , es decir, ocho en total. Eliminando las variables internas quedan dos ecuaciones del cuadripolo.

Despejando las variables del transistor en función de i_b y sustituyendo en las de salida y entrada, éstas quedan :

$$v_c=-Z_C\left(\frac{\alpha}{1-\alpha}i_b+i_s\right)$$

$$e=\frac{1}{39I_E(1-\alpha)}i_b+Z_E\frac{i_b}{1-\alpha}$$

Sustituyendo i_b de la segunda en la primera tenemos las ecuaciones técnicas del cuadripolo:

$$v_c=\frac{-\alpha Z_C}{\frac{1}{39I_E}+Z_E}e-Z_C i_s$$

$$i_b=\frac{1}{\frac{1}{39I_E(1-\alpha)}+\frac{Z_E}{1-\alpha}}e$$

De las ecuaciones se leen los parámetros técnicos:

$$A=\frac{-\alpha Z_C}{\frac{1}{39I_E}+Z_E}$$

$$Z_s = Z_C, \quad r=0$$

$$Z_{ec}=\frac{1}{39I_E(1-\alpha)}+\frac{Z_E}{1-\alpha}$$

8.8.- Amplificación.

De las ecuaciones leemos el parámetro amplificación sin carga,A (también llamado **ganancia**):

$$A=-\frac{\alpha Z_C}{\frac{1}{39I_E}+Z_E}$$

Esta fórmula da lugar a los siguientes comentarios:

Si las impedancias complejas son resistivas (o tienen fases pequeñas), la amplificación es **negativa**, la salida está en contrafase con la entrada.

Z_E es perjudicial para la amplificación. Si $Z_E = 0$, la amplificación es máxima y vale $A = -39I_E \alpha Z_C$. Si la impedancia de colector es resistiva $\alpha I_E R_C$ es aproximadamente la caída de tensión en la resistencia de carga. Si se eligió la tensión de colector igual a la mitad de la alimentación, la caída de tensión será también la mitad de la alimentación, con lo que la amplificación aproximada será $A_{max} = -20E_{BB}$.

Si, por el contrario, $1/39I_E << Z_E$, la amplificación se simplifica a (teniendo en cuenta que α es aproximadamente 1) $A = -Z_C / Z_E$. La amplificación queda no solamente independiente de la temperatura sino también de los parámetros del transistor, sólo depende de las impedancias del circuito. Se dice que Z_E estabiliza la amplificación, con el precio de disminuirla. Si, por ejemplo, se elige $Z_E = 0{,}1\ Z_C$, lo que es bastante normal, se tiene una amplificación -10, no muy alta pero perfectamente estabilizada.

8.9.- Impedancia de entrada.

La impedancia de entrada del transistor, sin incluir el circuito de polarización, vale

$$Z_{ec}=\frac{1}{39I_E(1-\alpha)}+\frac{Z_E}{1-\alpha}$$

y recordando que 1/ (1 - α) vale aproximadamente β (también llamada frecuentemente β = h_{21} por coincidir con ese parámetro híbrido del transistor) puede escribirse:

$$Z_{et}=\beta\left(\frac{1}{39 I_E}+Z_E\right)$$

En esta expresión vemos que Z_E es también favorable, reflejándose a la entrada multiplicado por β.

La impedancia de entrada así calculada es la del transistor, pero si queremos calcular la impedancia de entrada con una polarización como la de la figura 1, habrá que calcular el resultado de poner en paralelo con la entrada las resistencias R_1 y R_2. Es decir que la impedancia de entrada del circuito será:

$$Z_{ec} = Z_{et} \; // \; R_1 \; // \; R_2$$

designando con // el resultado de poner en paralelo.

8.10.- Parámetros de salida.

De las ecuaciones se lee directamente Z_s = Z_C. De esta forma se ve el inconveniente de aumentar la impedancia de colector, que aumenta la amplificación, pero también la impedancia de salida.

El parámetro técnico r de realimentación de corriente es cero. De ahí se deduce que las impedancias conectadas a la salida no se reflejan a la entrada, el transistor separa totalmente la entrada y la salida.

8.11.- Ecuaciones con impedancia de base.

La figura 8-6 representa el circuito de un amplificador con una impedancia conectada en la entrada de base (puede representar la impedancia del generador si éste no es ideal).

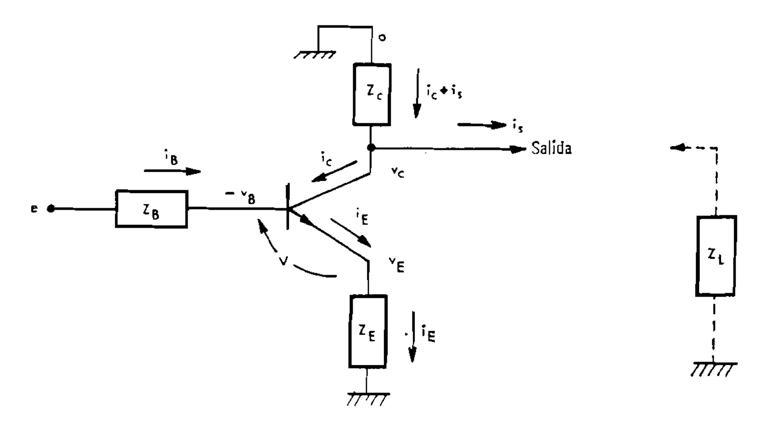


Fig. 8-6.

Las ecuaciones pasan a ser:

Transistor

$$i_E=39\,I_E v$$
$$i_C=\alpha i_E$$
$$i_B=(1-\alpha)\,i_E$$

Circuito

$$e=Z_B i_B+v+Z_E i_E$$
$$v_E=Z_E i_E$$
$$v_C=-Z_C(i_C+i_S)$$

Ecuación de carga

$$v_C=Z_L i_s$$

de donde, haciendo las mismas operaciones del párrafo 8.7, se obtienen los nuevos parámetros técnicos :

$$A=\frac{-\alpha Z_C}{\frac{1}{39}I_E+Z_E+(1-\alpha)\,Z_B}$$
$$Z_s=Z_C$$
$$Z_{ec}=\frac{1}{39\,I_E(1-\alpha)}+\frac{Z_E}{1-\alpha}+Z_B$$
$$r=0$$

De estas expresiones se deduce que el efecto de Z_B es perjudicial para la amplificación y se suma simplemente a la impedancia de entrada. No tiene el efecto estabilizador de Z_E porque en la fórmula de la amplificación aparece dividido por $1-\alpha$, y no es independiente de los parámetros del transistor.

8.12.- Circuito equivalente.

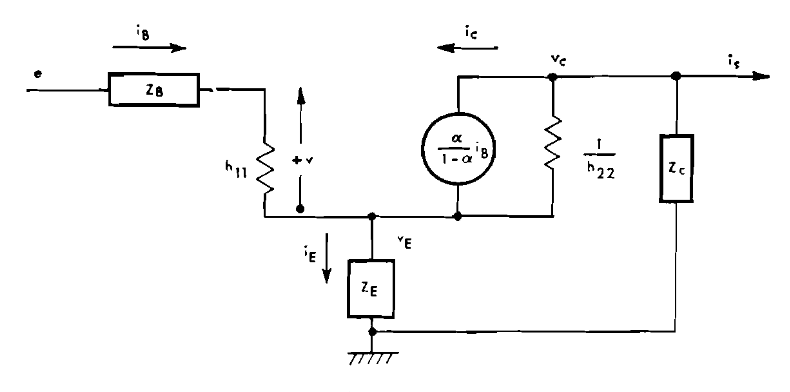


Fig. 8-7.

Si sustituimos el transistor por su circuito equivalente (usando los parámetros híbridos) se obtiene el circuito equivalente del amplificador, dibujado en la figura 8-7:

En este circuito, los parámetros híbridos del transistor valen, como se vio en el capítulo anterior:

$$h_{11}=\frac{1}{39\,I_E(1-\alpha)}$$
$$h_{21}=\frac{1}{(1-\alpha)}$$

El valor de h_{12} es despreciable. El valor de h_{22} es también muy pequeño, pero se tiene en cuenta fácilmente con una impedancia paralelo. A veces se incluye en Z_C (en paralelo), lo que es una aproximación suficientemente buena.

Teniendo en cuenta el valor de la impedancia de entrada calculada anteriormente, los parámetros híbridos **del circuito completo** serán ahora :

$$H_{11}=Z_B+\frac{1}{(1-\alpha)\,39\,I_E}+\frac{Z_E}{1-\alpha}$$
$$H_{12}=0$$
$$H_{21}=-\frac{\alpha}{1-\alpha}=-\beta$$
$$H_{22}=-\frac{1}{Z_C}$$

8.13.- Estabilización del punto de trabajo con la temperatura.

Los parámetros del transistor varían con la temperatura, con lo que el punto de trabajo no permanece en los valores de polarización calculados. La variación de los parámetros del transistor suele darse a través de dos coeficientes de variación, que son:

$$\left(\frac{\delta V_{BE}}{\delta t}\right)_{I_E=Cte.} \quad y \quad \frac{dI_{co}}{dt}$$

El primero expresa la variación de la tensión de base V_{BE}, en adelante **V**, con la temperatura para un valor de la intensidad de colector y vale aproximadamente - 2 mV / °C. El segundo da la variación de I_{co} con la temperatura y es, a su vez, muy variable, ya que el valor se triplica por cada 10°C de variación. El parámetro α no varía con la temperatura.

Calcularemos ahora la variación del punto de trabajo en función de la temperatura con polarización constante de la siguiente forma:

V (tensión de base emisor) es función de I_E y de la temperatura y podremos escribir:

$$dV=\left(\frac{\delta V}{\delta I_E}\right)_{t=Cte} dI_E+\left(\frac{\delta V}{\delta t}\right)_{I_E=Cte} dt$$

La primera derivada parcial puede calcularse y vale:

$$I_E=Ke^{39V}$$
$$V=\frac{1}{39}\log_e\frac{I_E}{K}$$
$$\left(\frac{\delta V}{\delta I_E}\right)_{t=Cte}=\frac{1}{39 I_E}$$

Diferenciando ahora la ecuación de entrada:

$$E=R_B I_B+R_E I_E+V$$
$$0=R_B dI_B+R_E dI_E+\frac{\delta V}{\delta t}dt+\frac{1}{39 I_E}dI_E$$

Diferenciando la ecuación de la intensidad de base y sustituyendo arriba, se obtiene:

$$I_B=(1-\alpha)I_E-I_{co}$$
$$R_B(1-\alpha)dI_E-R_B dI_{co}+R_E dI_E+\frac{\delta V}{\delta t}dt+\frac{1}{39 I_E}dI_E=0$$

de donde, dividiendo por dt y despejando dI_E /dt, se obtiene:

$$\frac{dI_E}{dt}=\frac{dI_{co}}{dt}\frac{R_B}{R_B(1-\alpha)+R_E\frac{1}{39 I_E}}-\frac{\delta V}{\delta t}\frac{1}{R_B(1-\alpha)+R_E+\frac{1}{39 I_E}}$$

y teniendo en cuenta que $I_C = \alpha I_E + I_{co}$:

$$\frac{dI_C}{dt}=\frac{dI_{co}}{dt}\left[\frac{\alpha R_B}{R_B(1-\alpha)+R_E+\frac{1}{39I_E}}+1\right]-\frac{\delta V_{BE}}{\delta t}\frac{\alpha}{R_B(1-\alpha)+R_E+\frac{1}{39I_E}}$$

y como $V_C = E_{BB} - R_C I_C$, es decir, $dV_C = - R_C dI_C$, se obtiene finalmente como solución del problema:

$$\frac{dV_C}{dt}=-\frac{dI_{co}}{dt}\left[\frac{\alpha R_B R_C}{R_B(1-\alpha)+R_E+\frac{1}{39I_E}}+R_C\right]+\frac{\delta V_{BE}}{\delta t}\frac{\alpha R_C}{R_B(1-\alpha)+R_E+\frac{1}{39I_E}}$$

Suponiendo $\alpha = 1$ y despreciando $1/39I_E$ frente a R_E resulta:

$$\frac{dV_C}{dt}=-\frac{dI_{co}}{dt}R_C\left[\frac{R_B}{R_E}+1\right]+\frac{\delta V}{\delta t}\frac{R_C}{R_E}$$

Como se ve, R_B es perjudicial, R_E **estabiliza**, por lo que conviene hacerlo grande (compatible con una amplificación suficiente). Para no perder amplificación a frecuencias altas, puede cortocircuitarse con un condensador.

Los dos términos son **negativos**, por lo que ambos efectos se suman, tendiendo a aumentar la intensidad de colector, es decir, a disminuir la tensión de salida. Parcialmente, puede compensarse este efecto haciendo que la polarización de entrada (E) disminuya con la temperatura, usando un divisor de tensión con resistencias sensibles a la temperatura, por ejemplo con una de tipo NTC a masa, que, al subir la temperatura, bajará la polarización de entrada, aumentando la tensión de salida a través de la amplificación.

8.14.- Seguidor de emisor.

Una utilización especial del transistor es el **seguidor de emisor**, representado en la figura 8-8.

Suponemos el circuito correctamente polarizado. Las ecuaciones para señales son :

$$i_e = 39I_E\, v$$
$$i_b = (1 - \alpha)\, i_e$$
$$e = v + v_e$$

Sustituyendo v de la primera en la tercera se obtiene la primera ecuación técnica:

$$v_e = e - \frac{1}{39 I_E} i_e$$

La segunda es la segunda ecuación técnica. Por consiguiente, los parámetros técnicos son:

$A = 1$
$Z_s = 1/\ 39I_E$
$Z_{et} = \infty$
$r = 1 - \alpha$

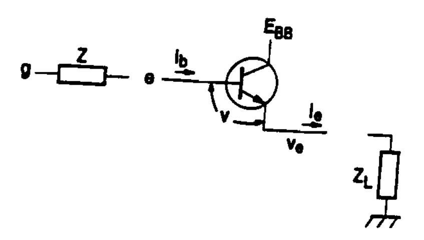


Fig. 8-8.

Obtenemos el resultado de que la amplificación es 1 (sin carga) con impedancia de salida baja. La impedancia conectada a la salida (la suponemos mayor que la interna de salida) se refleja a la entrada dividida por r, o sea, multiplicada por β. Esto justifica la aplicación del seguidor de emisor como **adaptador de impedancia.** No amplifica pero refleja la carga aumentada en un factor β.

Por último veamos como se refleja una impedancia Z conectada en serie a la entrada, simulando la impedancia del generador g, cuando no sea despreciable. La ecuación de entrada queda ahora:

$$g = Zi_b + v + v_e$$

y, sustituyendo la segunda del transistor:

$$g = Z\ (1 - \alpha) i_e + i_e\ /39I_E + v_e$$

con lo que se ve que la impedancia de salida ha quedado aumentada en la del generador dividida por β.

En resumen, la señal no es afectada, la impedancia de carga se refleja a la entrada multiplicada por β y la del generador se refleja a la salida dividida por β.

8.15.- Amplificador selectivo.

En general, en el circuito de la figura 8-9, se elige Z_E resistivo y, por lo tanto, independiente de la frecuencia, y la fórmula de la amplificación queda:

$$A=\frac{-\alpha Z_C}{\frac{1}{39 I_E}+Z_E}=\frac{-\alpha Z_C}{\frac{1}{39 I_E}+R_E}\approx\frac{-Z_C}{R_E}$$

tomando $\alpha = 1$ y despreciando $1/39I_E$ frente a R_E . En esta expresión se ve que la dependencia de la amplificación en función de la frecuencia es proporcional a la variación de la impedancia del colector respecto a la frecuencia.

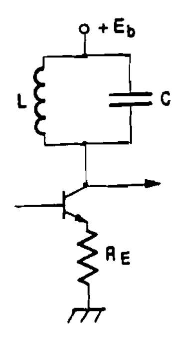


Fig. 8-9.

Si elegimos un circuito resonante como impedancia de colector, como se ve en la figura 8-9, la variación de la impedancia es la del circuito resonante paralelo y la amplificación sigue la misma curva, como se ve en la figura 8-10.

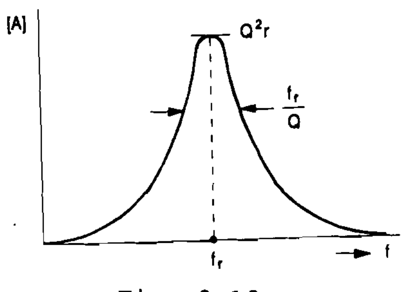


Fig. 8-10.

8-16.- Amplificador diferencial.

El amplificador de la figura 8-11 es un conjunto de dos amplificadores iguales cuyo emisor es común.

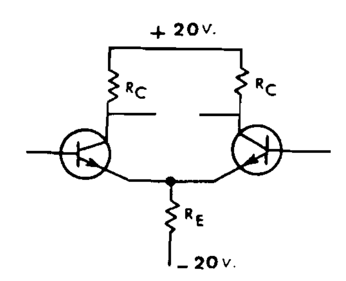


Fig.8-11.

Si llamamos con subíndice 1 a las señales del transistor de la izquierda y con subíndice 2 a las señales de la derecha, suponiendo entradas e_1 y e_2 en las bases y salidas v_{C1} y v_{C2} en los colectores, con una intensidad de salida i_s conectada entre los colectores, es decir que la carga se conecta **entre los colectores**, las ecuaciones del circuito para señales son:

$$i_{e1}=39\,I_E(e_1-v_e)$$
$$i_{e2}=39\,I_E(e_2-v_e)$$
$$i_{c1}=\alpha\, i_{e1}$$
$$i_{c2}=\alpha\, i_{e2}$$
$$i_{b1}=(1-\alpha)\,i_{e1}$$
$$i_{b2}=(1-\alpha)\,i_{e2}$$
$$v_{c1}=-(i_c+i_s)\,R_C$$
$$v_{c2}=-(i_c-i_s)\,R_C$$

Hemos supuesto los dos transistores idénticos (con los mismos parámetros) y polarizados en el mismo punto de trabajo (I_E igual para ambos, es decir, que la intensidad que circula por R_E es $2I_E$). Restando las ecuaciones dos a dos y eliminando las variables internas, resultan las ecuaciones técnicas tomando como variables las **diferencias** entre las variables simétricas:

$$v_{c1}-v_{c2}=-39\,I_E\alpha\,(e_1-e_2)-2R_C i_s$$
$$i_{b1}-i_{b2}=\frac{1-\alpha}{39\,I_E}\,(e_1-e_2)$$

Estas ecuaciones revelan que el **amplificador diferencial** se comporta como un amplificador de un transistor para las **diferencias de señales**, con la misma amplificación e impedancia de entrada, pero como si no hubiese resistencia de emisor. Es decir, se conserva la amplificación alta a pesar de que el punto de trabajo está estabilizado. La resistencia de salida es $2R_C$.

El amplificador diferencial amplifica la **diferencia** entre e_1 y e_2, obteniendo la salida **entre** los dos colectores. Si conecta

a tierra la entrada 1, el amplificador amplifica e_2 sin invertirla, y si se conecta a tierra la entrada 2, el amplificador amplifica la entrada e_1 **invertida**. Si las señales son iguales, $e_1 = e_2$ no hay señal de salida. El amplificador **no** amplifica las señales simétricas ni tampoco las perturbaciones o ruidos simétricos.

Un tipo especial de amplificador diferencial es el de **contrafase** o **push-pull** (figura 8-12):

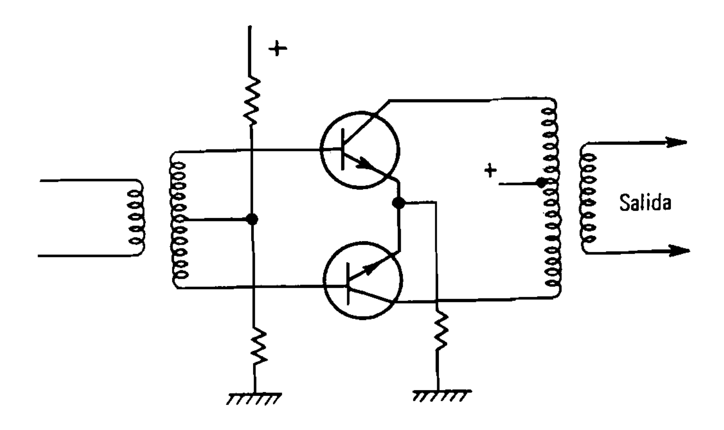


Fig. 8-12.

El transformador de entrada transforma la señal de entrada e en dos señales +e y -e, con diferencia de entradas en el amplificador diferencial 2e. La diferencia de salidas $v_{c1} - v_{c2}$ vuelve a transformarse en el transformador de salida dividiéndose por dos, con lo cual la amplificación queda en el mismo valor, pero con entrada y salida que ya no son diferenciales, por ejemplo, referidas a tierra. Este circuito tiene la ventaja adicional de que cuando la señal se amplifica por uno de los transistores por debajo del punto de trabajo, el otro transistor la amplifica tomándola por encima del punto de trabajo. Si la amplificación tiene alguna alinealidad (por ejemplo, la amplificación aumenta por encima del punto de trabajo y disminuye por debajo) se compensa la mayor amplificación de un lado con menor amplificación en el otro y la amplificación suma permanece aproximadamente constante. Esto quiere decir que se pueden usar señales algo mayores con la misma amplificación que con señales pequeñas. Naturalmente, el rango de frecuencias que pueden amplificarse con el push-pull está limitado por el transformador.

8-17.- Amplificadores de alta frecuencia.

Vamos a ver los fenómenos que se presentan al tratar de amplificar altas frecuencias y la forma de paliar su efecto.

En general, a altas frecuencias no es posible despreciar el efecto de las capacidades entre las diferentes partes del

transistor, y hay que operar teniendo en cuenta estas capacidades. En particular, la capacidad entre la base y el emisor puede tenerse en cuenta incluyendo un condensador equivalente en paralelo con la entrada y la capacidad entre el colector y tierra, a través del emisor o del circuito exterior, puede tenerse en cuenta con un condensador equivalente conectado a la salida. Estas capacidades se llaman **parásitas** y es necesario tenerlas en cuenta a partir de ciertos valores de la constante de tiempo que forman con las impedancias de entrada y salida, como se verá más adelante, cuando calculemos las frecuencias "de corte" de los amplificadores.

Más difícil es tener en cuenta la capacidad entre base y colector, o sea, entre entrada y salida. Esta capacidad puede sustituirse por una capacidad de entrada a través del **efecto Miller**, que veremos a continuación.

El efecto Miller se demuestra a la vista de la figura 8-13, en que el triángulo representa cualquier amplificador (en particular un transistor) de amplificación -A. Suponemos una impedancia de entrada Z_E y una capacidad interna entre entrada y salida C (o sea, entre base y colector), que representaremos con un condensador exterior.

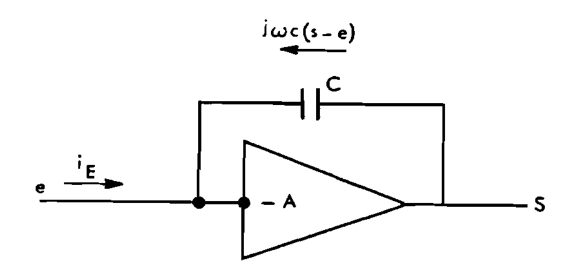


Fig. 8-13.

La intensidad de entrada valdrá:

$$i_E=\frac{e}{Z_E}-j\omega C(s-e)$$
$$y\ como\ s=-Ae$$
$$i_E=\frac{e}{Z_E}+j\omega C(A+1)e$$
$$\frac{i_E}{e}=\frac{1}{Z_E}+j\omega C(A+1)$$

Aparece la capacidad C(A+1) ≈ AC (suponiendo A>>1) en paralelo con la entrada. Se dice que C **se refleja** como AC a la entrada.

En el comportamiento del circuito a altas frecuencias hay

que evitar la influencia de la autoinducción y capacidad de las conexiones todo lo posible. Para ello es conveniente "separar" la polarización con autoinducciones o poner condensadores que bajen la impedancia de la alimentación a altas frecuencias. Por ejemplo, en la figura 8-14 el condensador C_1 garantiza que la impedancia en el punto de polarización A es cero y no pasa señal del amplificador a la polarización, mientras que L_1 hace que la polarización no cargue a la entrada (alta impedancia a altas frecuencias, sin perturbar la polarización de c.c.). L_2 es la carga del transistor y C_2 no deja pasar señales a otros circuitos a través de la alimentación del colector.

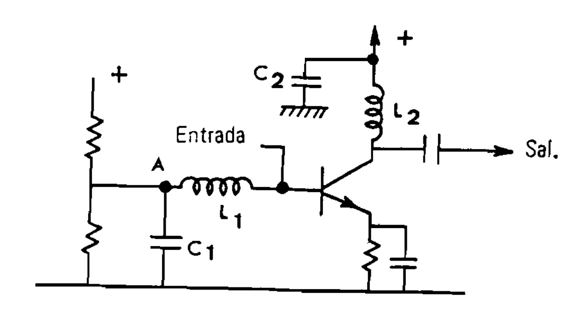


Fig. 8-14.

Todas las componentes deben ser cuidadosamente seleccionadas. Por ejemplo, las autoinducciones deben usarse por debajo de la frecuencia "propia" a partir de la cual la capacidad entre las espiras domina sobre la autoinducción, que se convierte así en un condensador.

8.18.- Inversor.

Un tipo especial de amplificador es el **inversor** de la figura 8-15.

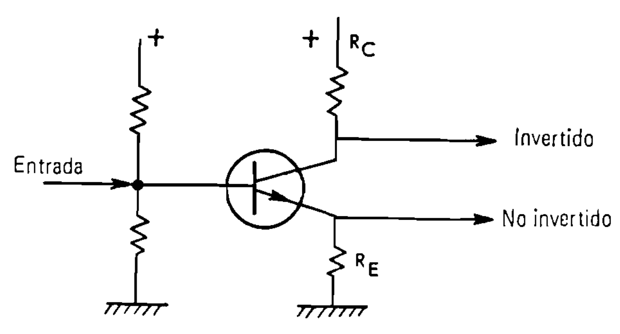


Fig. 8-15.

Las resistencias R_C y R_E se eligen iguales para obtener una amplificación -1 en la salida de colector y 1 en la de emisor.

Con ello se obtiene una señal invertida respecto a la entrada en la primera y no invertida en la segunda. Si se quiere aproximar más, las resistencias pueden hacerse ligeramente distintas para compensar el valor de $\alpha \neq 1$, haciendo la resistencia de colector ligeramente mayor, en proporción $1/\alpha$.

Con este circuito se convierte la señal de entrada en un par de señales simétricas, aptas, por ejemplo, para alimentar un amplificador diferencial. También se puede usar simplemente como inversor, utilizando solamente la salida en el colector.

8.19.- Acoplamiento entre etapas.

Los amplificadores suelen estar compuesto de un número de etapas sucesivas, cada una de las cuales amplifica las señales e influye sobre la respuesta de frecuencia. Pero, además, es necesario enlazar estas etapas entre sí, lo que se hace con el circuito de **acoplamiento**, que puede ser tan sencillo como la unión directa de una etapa con la siguiente, pero que, en todo caso, debe ser estudiado para ver su efecto sobre los circuitos anterior y siguiente.

El circuito de acoplamiento entre dos etapas cumple las siguientes misiones:

a) Quitar la polarización a la señal procedente de la etapa anterior y polarizar la siguiente, es decir, **cambiar** la polarización.

b) Dejar pasar las señales atenuándolas lo menos posible.

El acoplamiento está colocado en el camino de la señal e influye en la amplificación a través de dos efectos:

a) Cargando al circuito anterior como carga adicional o formando parte de su carga.

b) Con su propia respuesta de frecuencia que forma parte del conjunto.

Para calcular la amplificación total en función de la frecuencia en un amplificador de varias etapas se comienza el cálculo a la salida y se continúa hacia la entrada de la siguiente forma (figura 8-16):

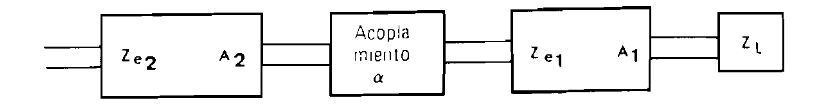


Fig. 8-16.

Se calcula la amplificación A_1 de la última etapa con su

carga Z_L y la impedancia de entrada Z_{e1}, ambas en función de la frecuencia en su caso.

Se calcula la amplificación α y la impedancia de entrada Z_α del circuito de acoplamiento, cargado con Z_{e1} (fig. 8-17).

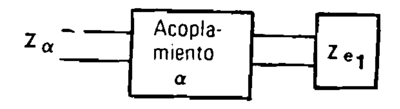


Fig. 8-17.

Se calcula la amplificación A_2 y la impedancia de entrada de la penúltima etapa, cargada con Z_α (fig. 8-18).

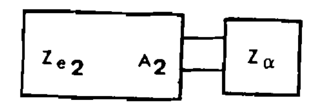


Fig. 8-18.

Así se continúa hacia la entrada si hay más etapas. La amplificación del conjunto será $A_1\ \alpha_1\ A_2\ \alpha_2 \ldots$ y la impedancia de entrada del amplificador será la última calculada, o sea, la de la primera etapa. La impedancia de salida del amplificador será la de la última etapa. Todas ellas funciones de la frecuencia en su caso.

Previamente se ha debido comprobar que los circuitos de acoplamiento cumplen su misión de cambiar la polarización.

8.20.- Tipos de acoplamiento y amplificadores.

En realidad la clasificación de los acoplamientos se corresponde con la de los amplificadores, ya que el acoplamiento se elige según va a ser el amplificador. Podemos citar las siguientes clases de acoplamiento, que no forman una clasificación completa, sino mas bien una enumeración de los más característicos:

a) Acoplamiento RC.

Es el más usado y típico de los **amplificadores de banda ancha**. Utiliza la propiedad del condensador de separar la c.c. de polarización y dejar pasar las señales, siempre que su frecuencia sea suficientemente alta.

b) Acoplamiento por transformador.

Utiliza la propiedad del transformador de transmitir las

señales de c.a. del primario al secundario y hacer independiente el nivel del primario y secundario. Es más caro que el anterior y tiene una respuesta de frecuencia limitada, pero permite sumar señales, invertirlas (como hemos visto en el push- pull o **amplificador de contrafase**) y separa muy bien los niveles de c.c. A pesar de la banda limitada de frecuencias puede considerarse de banda ancha.

c) Acoplamiento selectivo.

Utiliza un filtro como acoplamiento, consiguiendo así una determinada respuesta de frecuencia a través de su propia respuesta. Los **amplificadores selectivos** pueden serlo a través del acoplamiento o a través de la carga, como hemos visto con anterioridad.

d) Acoplamiento de c.c.

Si se quieren transmitir señales que consisten en desviaciones pequeñas, pero de variación lenta (las llamaremos de c.c.), con un espectro de frecuencias bajo, que no permite el uso de condensadores ni transformadores, hay que usar divisores de tensión que cambian las señales junto con la polarización, lo que presenta una dificultad especial. Los **amplificadores de c.c.** están caracterizados por este tipo de acoplamiento.

Además de los tipos de amplificadores citados aquí, podemos recordar otros tipos vistos con anterioridad, como el **amplificador diferencial** o restador y el **seguidor de emisor**, que es un adaptador de impedancias.

8.21.- Acoplamiento RC.

Consiste en separar las etapas con un condensador que separa la polarización y deja pasar las señales. El condensador forma un RC con la resistencia de entrada de la etapa que le sigue, de donde recibe su denominación. Su estudio podría remitirse al caso general del párrafo 8.16, pero por su interés vamos a hacerlo con más detalle.

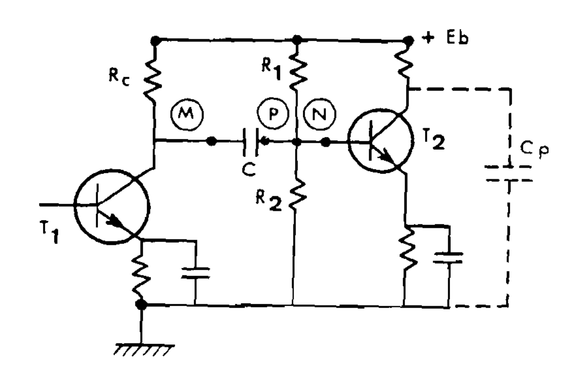


Fig. 8-19.

La figura 8-19 representa dos etapas de amplificación de transistores acopladas por RC a través del condensador C que las une. Para calcular la amplificación se corta por M y N, en la forma indicada en la figura 8-20. Primero se calcula la amplificación y la impedancia de carga del transistor T_2 sin circuito de polarización A_2 y Z_E. Después se calcula la amplificación (atenuación) α en el circuito de acoplamiento, incluyendo las resistencias R_1, R_2 y Z_E (que también es resistiva). A continuación se calcula la amplificación de T_1, A_1 con la carga formada por C, R_1,R_2 y Z_E. La amplificación total es $A_1\ \alpha\ A_2$, la impedancia de entrada es la de T_1 y la de salida es la de T_2.

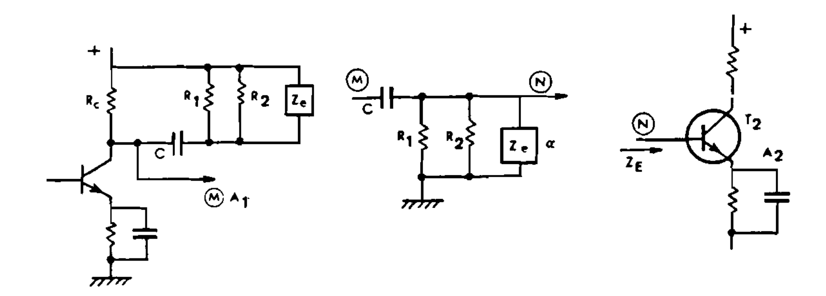


Fig. 8-20.

El condensador forma un RC como hemos visto entre M y N, de forma que α es la respuesta de frecuencia de un RC, con **frecuencia de corte** dada por:

$$f_C=\frac{\omega_C}{2\pi}=\frac{1}{2\pi R_e C}$$

siendo C el condensador de acoplamiento y R_e la impedancia de entrada de la etapa siguiente (con las resistencias de polarización incluidas). Si suponemos que el RC determina el corte a bajas frecuencias (más adelante detallaremos este extremo), la curva total tendrá la forma de una respuesta RC. La respuesta continuaría alta de forma indefinida si no se considerasen las capacidades parásitas, representadas en la figura 8-19 por el condensador C_p, que cortocircuita la salida a partir de cierta frecuencia. El conjunto, con una respuesta como la de la figura 8-21, es un amplificador de **banda ancha**.

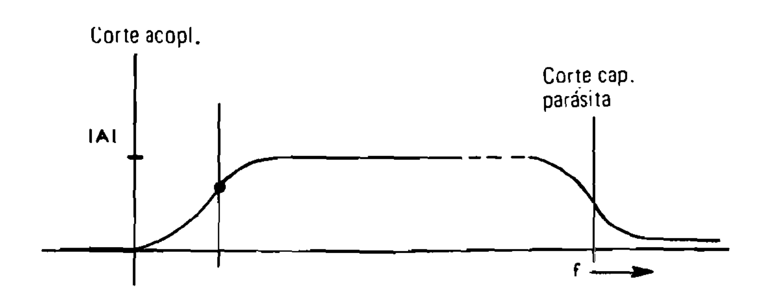


Fig. 8-21.

8.22.- Acoplamientos por transformador y selectivo.

El transformador es un buen elemento de acoplamiento. Se usa como se indica en la figura 8-22. Se usa un transformador fuertemente acoplado que es un elemento de banda ancha y refleja la impedancia del secundario dividida por n^2 (n relación de transformación). Las polarizaciones de primario y secundario son independientes. La respuesta de frecuencia está limitada por abajo por la autoinducción finita y por arriba por la capacidad distribuida del transformador.

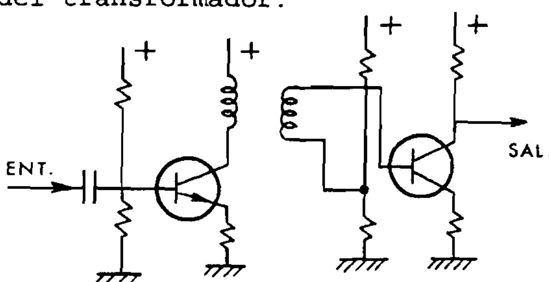


Fig. 8-22.

También se puede usar un acoplamiento como filtro, para obtener respuestas selectivas. Por ejemplo, la figura 8-23 representa un acoplamiento con filtro de transformador doblemente sintonizado y curva selectiva ensanchada, como hemos estudiado con anterioridad.

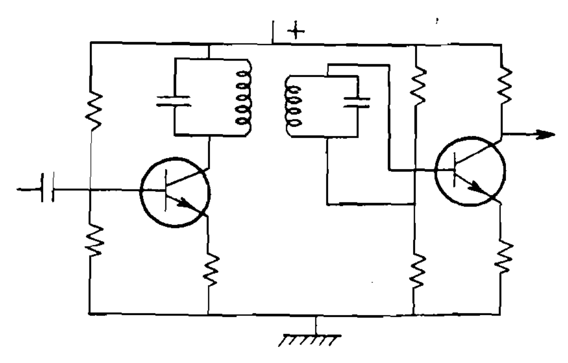


Fig. 8-23.

La figura 8-24 ilustra el acoplamiento con un filtro RLC, cuya respuesta de frecuencia será selectiva, obteniendo resultados parecidos a los del amplificador selectivo del párrafo 8.12.

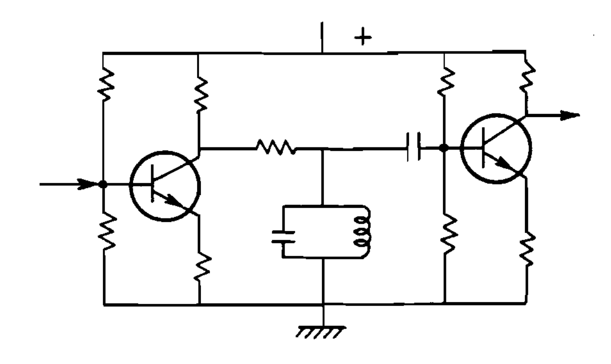


Fig. 8-24.

8.23.- Amplificador de c.c.

Aunque se acople a través de condensadores muy grandes, no es posible amplificar señales de c.c., es decir, desviaciones pequeñas pero no oscilatorias y que no pasan por el condensador. El acoplamiento de c.c. ha de hacerse sólo con resistencias, como se ve en la figura 8-25.

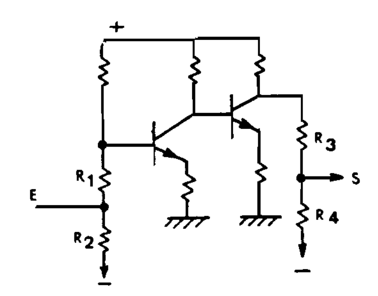


Fig. 8-25.

La dificultad del cambio de polarización con divisores de tensión hace necesaria una alimentación auxiliar negativa, porque si los divisores se limitan a rebajar el nivel de polarización, se rebaja también el nivel de señal. Los divisores de tensión deben cumplir la condición de cambio de polarización y no atenuar demasiado las señales.

8.24.- Amplificadores de banda ancha. Niveles.

Los amplificadores de banda ancha están proyectados para amplificar una extensa gama de frecuencias. El acoplamiento suele ser del tipo RC ya descrito, aunque excepcionalmente puede ser de c.c.

Lo típico de la respuesta de frecuencia de estos amplificadores es la presencia de extensos intervalos de frecuencia en los que la amplificación es constante, llamados **niveles**, separados por zonas de transición producidas por los circuitos RC presentes. Tendremos en cuenta que la curva de respuesta de frecuencia de un circuito RC puede aproximarse, como se ve en la figura 8-26 por dos largos tramos constantes que corresponden a las aproximaciones de condensador abierto a bajas frecuencias y condensador cortocircuitado a altas frecuencias, separadas por una zona de transición que corresponden a la frecuencia de corte del RC, que vale $1/2\pi RC$. Por encima de esta frecuencia, la impedancia del condensador es $1/\ 2\ \pi\ f\ C << R$ y por debajo es $R << 1/\ 2\ \pi\ f\ C$.

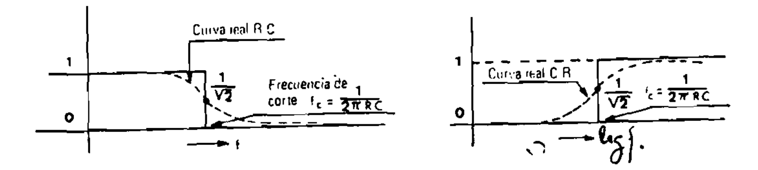


Fig. 8-26.

Vamos a tratar un ejemplo de amplificador de banda ancha, que es el dibujado en la figura 8-27. Dibujaremos la respuesta de frecuencia empezando por las frecuencias bajas (fig. 8-28).

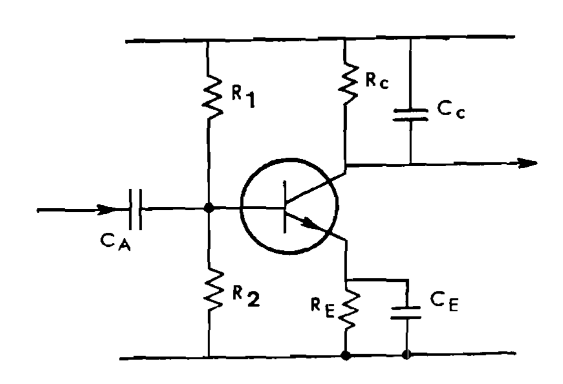


Fig. 8.27.

En este circuito habrá tres zonas de transición o zonas en las que cada uno de los tres condensadores, C_A, C_E y C_C, puede considerarse, con buena aproximación, que pasa de ser un circuito abierto a ser un cortocircuito. En primer lugar hay que calcular las resistencias "asociadas" a cada uno de ellos, en paralelo o en serie. Por el momento, consideraremos que las frecuencias de corte se presentan en el orden citado.

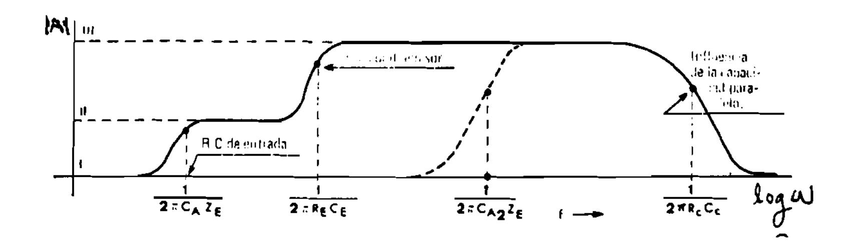


Fig. 8.28.

El condensador C_A forma un circuito CR con la resistencia de entrada del circuito del transistor. Sea R_e esta resistencia de entrada, es decir, el resultado de conectar en paralelo la resistencia de entrada del transistor (calculada con C_E abierto, es decir, con resistencia de emisor) con las del divisor de tensión R_1 y R_2. La constante de tiempo será R_e C_A, y el valor de la frecuencia de corte será 1/ 2 π R_E C_A y por debajo de este valor, el condensador se considera abierto y la amplificación es 0. Por encima de este valor, el condensador está cortocircuitado y la entrada ocurre directamente al transistor, como se ve en la figura 8-29. Suponemos que las demás transiciones **están alejadas** y no influyen entre sí. Por consiguiente, los otros dos condensadores se consideran abiertos, y el **circuito equivalente para el primer nivel** de la figura 8-28 es el de la figura 8.29.

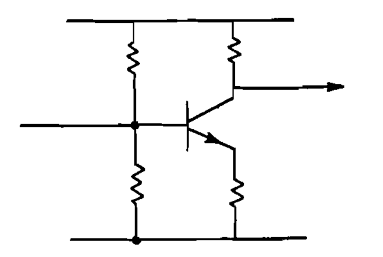


Fig. 10-16.

Fig. 8-29.

La amplificación de este nivel, calculado con el circuito de la figura 8-29, es válida hasta la transición siguiente, en que el condensador de emisor actúe como un cortocircuito. Calcularemos ahora la frecuencia de corte del emisor, que será 1/ 2 π R_E C_E. A partir de ese valor el condensador C_A se considera abierto, lo que aumenta la amplificación. El nuevo **circuito equivalente para el segundo nivel** es el de la figura 8-30. Ahora se calcula la nueva amplificación (sin resistencia de emisor) y esta constituye el segundo nivel (el más alto) de la figura 8-28.

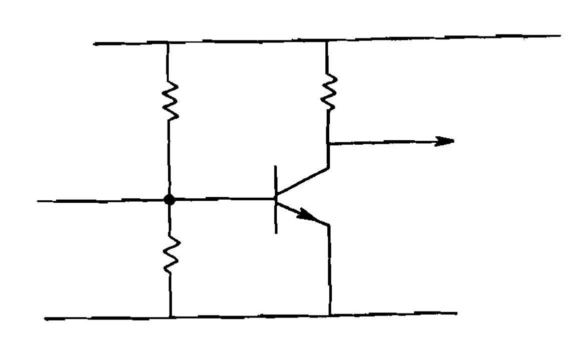


Fig. 8-30.

La última transición ocurre cuando la frecuencia se acerca a la de corte del circuito de colector, es decir $1/\ 2\ \pi\ R_C\ C_C$. A partir de ese valor, la salida está cortocircuitada por el condensador, y la amplificación es cero. El condensador C_C puede representar la capacidad parásita y ser un valor equivalente. A veces se pone intencionadamente para que el corte final a frecuencias muy altas esté bien determinado.

Los niveles así definidos no aparecen necesariamente todos. Por ejemplo, si el **orden** no es el indicado sino que $R_BC_B >> C_{A2}Z_e >> R_CC_C$ (con un menor condensador de acoplamiento) la curva de amplificación quedaría como se ve en la figura 8-28 de trazos. La amplificación cero hasta la transición de entrada oculta la transición de emisor.

Naturalmente, el amplificador se va a usar sólo en el nivel más alto de la figura 8-28, por lo que este intervalo es mucho mayor en la realidad que el representado en la figura, puede tener una extensión, por ejemplo, de 100 Hz. a 50.000 Hz. en un amplificador de audio, con una "dinámica" de 1:500, término que se usa algunas veces.

8.25.- Características generales de los amplificadores.

El comportamiento global de un amplificador está definido por sus actuaciones características, algunas de las cuales detallamos a continuación:

a) Amplificación o ganancia, impedancia de entrada y salida.

Estos tres parámetros técnicos tienen una definición muy precisa, que corresponde a la de los parámetros del cuadripolo y son función de la frecuencia.

En la práctica se suelen definir los tres a una frecuencia determinada . En los amplificadores de banda ancha se suele tomar una frecuencia central en la zona de utilización máxima. Los parámetros se conservarán en un amplio intervalo de frecuencias. En general la **amplificación** se suele dar con una carga

determinada, que podemos llamar **carga nominal** y, por lo tanto, no coincide con el parámetro técnico, que sería sin carga. **Ganancia** es lo mismo que amplificación, pero suele darse en **decibelios**. Recuérdese que el valor en dB. es 20 veces el logaritmo decimal de la amplificación. Así, la amplificación 100, por ejemplo, corresponde a 40 decibelios.

En los amplificadores selectivos, se definirán a la frecuencia de resonancia (la de amplificación máxima). En los amplificadores de c.c. se toma la amplificación de c.c., asimilada a frecuencia cero.

Las impedancias de entrada y salida son los parámetros técnicos ya definidos, a la frecuencia de referencia. En los amplificadores, casi sin excepción, el cuarto parámetro técnico es cero, y no hay influencia entre la salida y la entrada, no siendo entonces necesario definir las condiciones de medida de las impedancias de entrada y salida.

b) **Respuesta de frecuencia. Ancho de banda.**

La respuesta de frecuencia es la curva de amplificación (módulo) en función de la frecuencia.

A veces se da en escala semilogarítmica, ordenadas en decibelios y abscisas en frecuencia. También puede darse en escala logarítmica, ordenadas en decibelios y abscisas representadas en logaritmo decimal aunque se escriben las frecuencias. En ese caso, en abscisas aparecen las frecuencias potencias de 10 linealmente espaciadas. Con ello se representan mejor los intervalos enormes de los amplificadores de banda ancha. También algunas veces a duplicar la frecuencia se le llama "subir una octava", denominación más usada en control, como veremos.

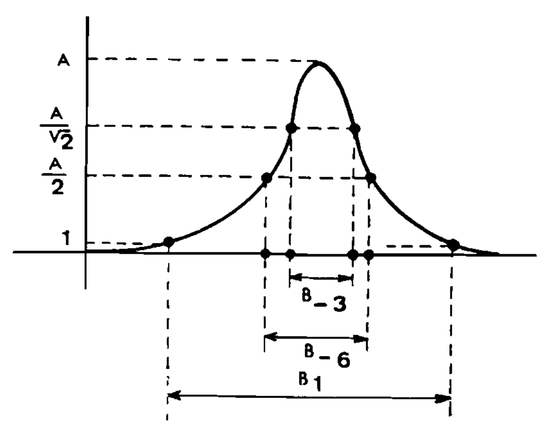


Fig. 8-31.

Se llama **ancho de banda** a la distancia en frecuencia entre dos puntos cuya amplificación es una fracción determinada del máximo. El ancho de banda presupone que la curva tiene una forma convencional con un solo máximo. Por ejemplo, en las curvas de

las figuras 8-31 (amplificador selectivo) y 8-32 (amplificador de c.c.) se definen los anchos de banda B_{-3} a $1/\sqrt{2}$ (a - 3 dB) o B_{-6} a 1/2 (a -6 dB) del máximo. El primero, a - 3 dB del máximo, es el más usado.

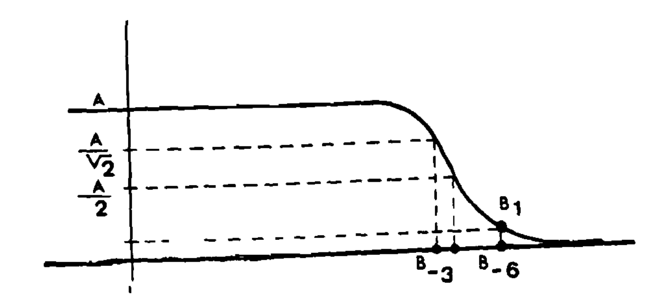


Fig. 8-32.

c) **Número de ruido.**

En el capítulo de señales hemos estudiado la definición del número de ruido de un sistema, que es directamente aplicable a los amplificadores. Más adelante veremos que el número de ruido de un amplificador es prácticamente el de su primera o primeras etapas, por analogía con los receptores.

d) **Tiempo de subida.** (Rise time).

En los amplificadores de banda ancha y continua es importante el tiempo de subida, que se define como el tiempo que la salida tarda en llegar a $1/\sqrt{2}$ del valor final cuando se aplica a la entrada un escalón unidad (fig. 8-33).

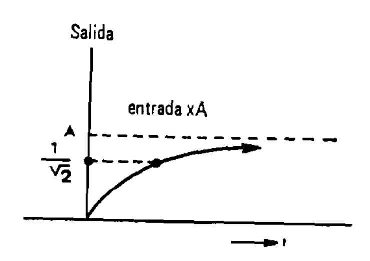


Fig. 8-33

En la mayoría de los amplificadores, la respuesta de frecuencia puede asimilarse aproximadamente a un circuito RC, que es el RC de corte de las altas frecuencias. Como el transitorio de un circuito RC es bien conocido, la salida, para entrada escalón unidad será (fig. 8-33):

$$s=A(1-e^{-\frac{t}{RC}})=A(1-e^{-2\pi f_c t})$$

teniendo en cuenta que la frecuencia de corte del amplificador, que prácticamente coincide con el ancho de banda en estos tipos de amplificadores, es $f_c = 1 / (2 \pi RC)$.

El tiempo que tarde la función s en llegar al valor $A/\sqrt{2}$ se calculará fácilmente :

$$\frac{1}{\sqrt{2}}=1-e^{-2\pi f_c t}$$

$$t=-\frac{1}{2\pi f_c}\ln\left[1-\frac{1}{\sqrt{2}}\right]=\frac{0,195}{f_c}$$

Esta fórmula da el tiempo de subida en función de la frecuencia de corte. En los amplificadores de c.c., la frecuencia de corte es igual al ancho de banda y en los amplificadores de banda ancha también aproximadamente, ya que la subida de la amplificación a bajas frecuencias es muy rápida y se mantiene prácticamente constante durante un gran intervalo hasta bajar a la frecuencia de corte.

Este parámetro tiene un gran interés en instrumentación, osciloscopios, etc..., como medida de la rapidez de reacción del sistema de medida.

e) **Potencia de salida.**

Si la amplitud de salida máxima es $S_{max.}$ y su resistencia de salida es R, la amplitud de salida varía con la carga, cumpliéndose la ecuación del generador:

$$S = S_{max.} - Ri$$

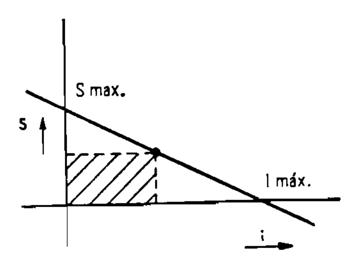


Fig. 8-34.

Esta recta (figura 8-34) da las variables de salida para distintos valores de la carga. La potencia de salida es el

producto Si, representado por el área rayada y que es máxima cuando el área sea máxima, es decir, en el punto central, cuando la resistencia exterior sea igual a la interior. La potencia en ese caso será $S_{max.2}$ / 4R.

Sin embargo, normalmente está limitada la potencia de salida por motivos de proyecto del amplificador, y no se debe sobrepasar la **potencia nominal** dada, que puede serlo directamente o dando el valor mínimo de la resistencia de carga.

f) **Distorsión.**

Los amplificadores reales no son totalmente lineales. Al aplicar una señal sinusoidal a la entrada, aparece una salida deformada, como se indica en la figura 8-35.

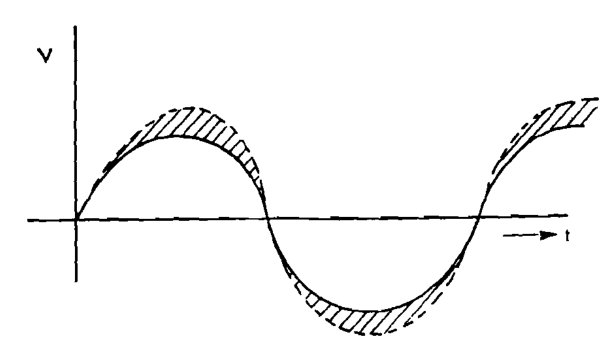


Fig. 8-35.

La deformación puede cuantificarse por la aparición de armónicos que vienen a sumarse a la onda fundamental del desarrollo de la onda de salida. La **distorsión** se mide por la fórmula:

$$d=\frac{V_{ef.salida}-V_{ef.fundamental}}{V_{ef.salida}}$$

que puede darse como coeficiente **d** o en tanto por ciento, **d:100**. En otro capítulo trataremos este tema dentro de la alinealidad de los amplificadores.

g) **Desviación de cero.** (drift).

Es un parámetro característico de los amplificadores de c.c. Se define como el valor que hay que dar a la entrada para dar cero a la salida. La desviación de cero no tendría importancia si se conservase constante, pero varía con el tiempo por alteraciones de las componentes y con la temperatura por calentamiento y cambio de valor de algunos componentes. Estos valores suelen darse en forma de **deriva máxima por hora** y **deriva máxima por grado** respectivamente. Los valores típicos son del

orden del milivoltio por hora y de unos microvoltios por grado. La desviación de cero es una forma de ruido de entrada para estos amplificadores.

CAPITULO 9

REALIMENTACION Y OSCILADORES

9.1.- Introducción.

Se entiende por realimentar el proceso de transformar la salida de un circuito o sistema y aplicarlo de nuevo a la entrada. El concepto tiene un campo de aplicación mucho más extenso que el de la Electrónica, especialmente en Cibernética y Automática. La figura 9-1 es un diagrama de bloques que muestra el principio de la realimentación : La cadena de realimentación da información sobre la salida y la mezcla (en general, la compara) con la entrada exterior. El sistema usa esta información para "corregir la salida", como veremos después. En el caso más simple, el circuito de realimentación toma una muestra que debe ser igual a la entrada y la resta de ésta. Si la diferencia no es cero, la "señal de error" se usa para corregir la salida.

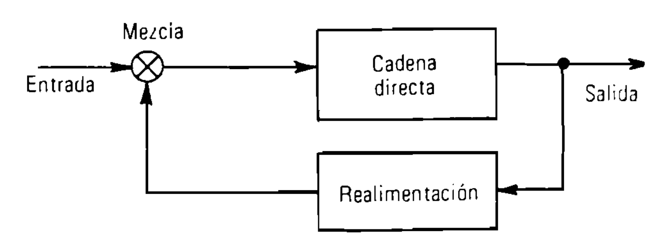


Fig. 9-1.

La realimentación ("**feed-back**") tiene también aplicación para producir osciladores, provocando una inestabilidad del circuito completo.

Por último, la gran precisión que se consigue con los amplificadores operacionales realimentados, permite usarlos como elementos de cálculo, o bien aislados, o bien formando parte de

un sistema de cálculo analógico.

Por último puede señalarse que la realimentación no deseada puede dar lugar a inestabilidades que es preciso evitar.

En este capítulo estudiaremos las aplicaciones electrónicas de la realimentación, en general y con amplificadores operacionales. También estudiaremos las condiciones de oscilación y los circuitos osciladores y los principios de cálculo analógico.

9.2.- Realimentación.

Comenzaremos estudiando el sistema de la figura 9-2. Un amplificador de amplificación A, impedancia de entrada infinito e impedancia de salida despreciable tiene una señal de salida s y su entrada recibe la suma de la entrada e y la fracción β de la salida que se suma en serie, por ejemplo, con un transformador u otro circuito análogo.

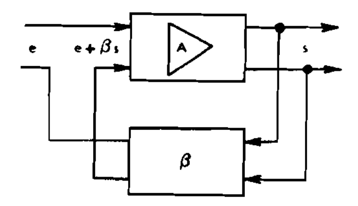


Fig. 9-2.

La entrada del amplificador será e + βs, y el amplificador la multiplicará por A, obteniendo:

$$s = A\ (e + \beta s).$$

De esta ecuación se despeja fácilmente la amplificación del sistema realimentado, que será:

$$\frac{s}{e} = \frac{A}{1-\beta A}$$

Esta es la fórmula general de la realimentación, en la que puede haberse usado cualquier recurso de cálculo operacional, en particular la transformación $j\omega$. El factor βA se llama **factor de realimentación**.

La fórmula de la realimentación permite enumerar ahora los casos interesantes:

a) Si (1 - βA) < 1, la amplificación con realimentación es **mayor** que sin ella. Se dice que hay una realimentación **positiva**. Normalmente se obtiene este efecto con **βA positivo real y menor**

que 1. Se usa muy pocas veces debido al peligro de inestabilidad que se presenta cuando el producto βA se acerca a 1, como veremos. No hay que olvidar que este efecto puede presentarse a cualquier frecuencia, es decir, que puede ocurrir que el factor de alimentación tenga un valor aceptable a una frecuencia y tome un valor peligroso o inestable a otra.

b) Si βA = 1 para alguna frecuencia, la amplificación se hace infinito, lo que quiere decir que, a esa frecuencia, para la que βA = 1, existe una señal de salida (de esa frecuencia) sin entrada. El sistema se convierte en un **generador** u **oscilador** sinusoidal. Al hacer e=0 en la figura 9-2, el sistema es un **amplificador que se alimenta a sí mismo.** La **condición de oscilación** es **βA(ω) = 1**, ecuación cuya solución es la frecuencia o pulsación de oscilación.

Si βA se hace mayor que 1, la ecuación de realimentación no refleja la solución real, que es divergente. Esto es debido a que el uso de la transformación compleja supone que la solución es sinusoidal y no deja aparecer soluciones que no lo son. En el párrafo siguiente estudiaremos con rigor el problema de la estabilidad, pero parece lógico que al sobrepasar la realimentación positiva necesaria para la oscilación, aparezcan oscilaciones divergentes, que aumentan de amplitud hasta que, debido a la saturación del amplificador, disminuye la amplificación y el factor de realimentación se hace unidad.

c) Si βA es muy grande en valor absoluto (negativo por razones de estabilidad), puede despreciarse el 1 en el denominador y la amplificación se convierte en :

$$\frac{s}{e}=-\frac{1}{\beta}$$

La amplificación **depende sólo de β**, se hace independiente de la amplificación A. En general, este caso, que se llama, de **realimentación negativa**, se obtiene con **A negativo y de valor absoluto muy grande**, ya que β suele ser menor que 1.

Este es el caso más interesante de realimentación. La no dependencia de A, con componentes diversos activos y pasivos, con parámetros que dependen de la temperatura, es fundamental. La realimentación β suele depender sólo de algunos componentes pasivos, que son estables y pueden ser de gran precisión. Se dice que la realimentación negativa **estabiliza** la amplificación.

9.2.- Estabilidad de los sistemas realimentados.

Las hipótesis de linealidad del párrafo anterior dan lugar a soluciones erróneas en la interpretación de la fórmula de realimentación. En efecto, al utilizar el operador jω se supone que todas las soluciones son estacionarias y lineales, descartando con ello otras de tipo transitorio convergente (sin interés por el momento) o **divergentes** que producen

inestabilidades.

Para considerar los fenómenos transitorios consideraremos el segundo miembro de la fórmula

$$\frac{s}{e}=\frac{A}{1-\beta A}$$

no como función de ω sino como función de transferencia entre dos transformadas de Laplace, función de **s**.

Como se recordará de la teoría de ecuaciones diferenciales o de la transformada de Laplace, los transitorios del sistema quedan definidos en función de las raíces del polinomio $1 - \beta A$ (que equivale a la ecuación característica de la ecuación diferencial que liga a s con e). Cada raíz r da lugar a una solución exponencial del tipo Ke^{jrt} (reales o complejos conjugados).

Este polinomio, función de **s**, tiene cinco clases de raíces que dan lugar a los correspondientes tipos de transitorios:

a) Raíces reales negativas, que dan lugar a transitorios exponenciales que tienden a cero y desaparecen al cabo de algún tiempo.

b) Raíces reales positivas, que dan lugar a soluciones exponenciales **divergentes**, que llevan al circuito a la saturación.

c) Raíces complejas conjugadas de parte real negativa, de la forma $-\alpha \pm \omega j$, que dan lugar a soluciones oscilatorias de pulsación ω , amortiguadas exponencialmente y que desaparecen al cabo de algún tiempo.

d) Raíces complejas conjugadas de parte real positiva, de la forma $\alpha \pm \omega j$,que dan lugar a soluciones oscilatorias exponencialmente **divergentes**, que crecerán hasta la saturación del amplificador, convirtiéndose entonces en oscilaciones rectangulares.

e) Raíces imaginarias conjugadas puras, que dan lugar a oscilaciones sostenidas y que estudiaremos en **osciladores**.

De las consideraciones anteriores se deduce inmediatamente la condición para que sólo haya soluciones convergentes, que es:

La condición necesaria y suficiente para que un sistema realimentado sea estable es que las raíces del denominador de su función de transferencia (los polos de la función de transferencia) sean negativas o de parte real negativa.

Aunque nos hemos referido hasta ahora a sistemas

realimentados **intencionadamente**, hay que hacer notar que en los sistemas reales se presentan muchas formas de acoplamiento no deseado y a veces poco detectable, por lo que también pueden presentarse formas de inestabilidad en amplificadores aparentemente sin realimentación. Estas realimentaciones son debidas a acoplamientos entre la salida y la entrada por inducción o capacidad parásitas y es un efecto a tener en cuenta en el proyecto de amplificadores.

Existen diversos **criterios de estabilidad** que permiten comprobar que se cumple la condición de estabilidad sin resolver la ecuación característica, por ejemplo, el **criterio de Routh**, que es un criterio de estabilidad derivado de los coeficientes de la ecuación. Sin embargo, es necesario conocer la ecuación y es muy opaco respecto a la influencia de los coeficientes en el resultado, lo que lo hace poco práctico en la realidad. Mucho más interesante son los criterios **basados en la respuesta de frecuencia**, que puede medirse directamente (incluso en el caso de inestabilidad puede medirse por partes) y no es necesario conocer exactamente la ecuación característica.

El método más usado y que estudiaremos con detalle en un capítulo posterior, dedicado al control, es el de **criterio de Nyquist**. El criterio de Nyquist se aplica a la vista del **diagrama de Nyquist**, que es la respuesta de frecuencia dibujada en coordenadas polares (amplitud y fase) o la respuesta compleja en función de ω . Para aplicar el criterio, dibujaremos el diagrama de Nyquist de la función de realimentación βA. Hay que hacer notar que en control se usa el diagrama de Nyquist de la función - βA, incluyendo el signo menos característico de los sistemas de control, pero aquí usaremos directamente la función de realimentación con signo positivo. Supondremos que el sistema es estable sin realimentación, lo que es normal pero no absolutamente necesario en algunos sistemas, como veremos en control.

El criterio de Nyquist dice así (fig. 9-3):

La condición necesaria y suficiente para que un sistema sea estable es que su diagrama de Nyquist en circuito abierto (βA) no rodee al punto crítico s = +1 del plano complejo, al variar ω entre $-\infty$ y $+\infty$.

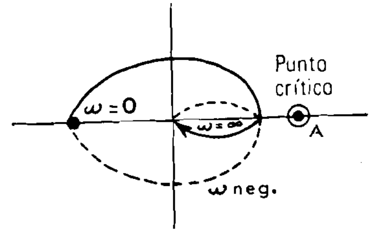


Fig. 9-3.

El diagrama de Nyquist da gran cantidad de información sobre el circuito realimentado, a partir del diagrama en circuito abierto (sin realimentar), información que permite estimar las

frecuencias propias de oscilación (incluso convergentes) y también corregir la curva de forma que no se produzca inestabilidad. En el capítulo de control se profundizará **más** sobre la aplicación del criterio de Nyquist.

9.3.- Osciladores.

Un oscilador es un amplificador con realimentación positiva unitaria, es decir, que cumple la ecuación $\beta A = 1$ para un valor determinado de ω . Puede decirse que un oscilador es un amplificador que se alimenta a sí mismo, sin entrada exterior (fig. 9-4).

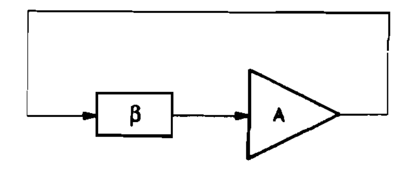


Fig. 9-4.

En teoría el oscilador es un caso límite entre oscilaciones convergentes y divergentes, pero en la práctica se utiliza la no linealidad del amplificador para **estabilizar las oscilaciones.**

En efecto, supongamos que la amplificación A de la figura es dependiente de la amplitud **a** de la señal de pulsación ω, normalmente decreciente con a, a causa de que el amplificador se aproxima a sus valores de saturación con amplitudes grandes, disminuyendo su amplificación. La ecuación o **condición de oscilación** $\beta A = 1$ es una ecuación compleja que se divide en dos ecuaciones (en fase y amplitud) con dos incógnitas, ω y **a**, lo que permite calcular la frecuencia o pulsación de las oscilaciones y su amplitud. Normalmente, A sólo depende de a en una amplia banda y la fase no depende de a, lo que permite separar la resolución de las ecuaciones en dos partes:

1) Se calcula la frecuencia **f** o la pulsación $\omega = 2 \pi f$ de la ecuación **fase** $(\beta A) = 0$.

2) Se calcula la amplitud de la ecuación $A(a)\beta(a,\omega) = 1$.

Normalmente, el proyecto del oscilador lleva consigo el proceso contrario:

1) Se proyecta β para que su fase sea cero a la pulsación ω deseada.

2) Se proyecta A para que satisfaga a $\beta A = 1$ a la pulsación y amplitud deseadas y sea $\beta A>1$ para valores menores de a y $\beta A<1$ para valores mayores. Con ello aparecerán oscilaciones divergentes de poca amplitud, que irá creciendo (y haciendo decrecer A) hasta satisfacer la condición de oscilación. Si sobrepasan la amplitud calculada, se hacen convergentes de nuevo, para tender siempre al valor que satisface exactamente a la

ecuación.

Los diferentes tipos de osciladores pueden tipificarse por el circuito de realimentación.

9.4.- Osciladores RC.

Cuando el circuito de realimentación está formado por resistencias y condensadores, el oscilador se denomina **RC**. Los casos más importantes son el de divisor RC y el oscilador de Wienn.

El oscilador más simple es el representado en la figura 9-5.

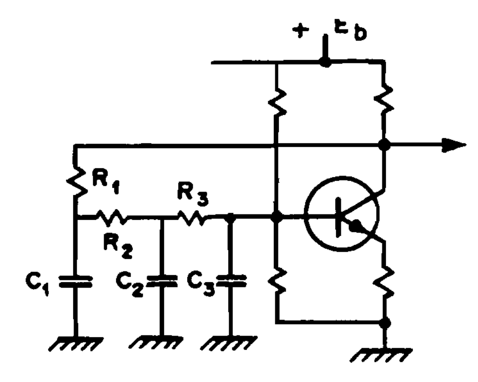


Fig. 9-5.

Cada divisor RC se proyecta para un desfase de 60º (para un desfase mayor la amplificación caería demasiado) y de forma que cada uno no cargue apreciablemente al anterior, lo que hace un total de 180º de desfase para β. La amplificación de cada divisor RC (para 60º de desfase) es 0,5, lo que da un valor de 1/8 al módulo de beta. El amplificador debe invertir la fase y amplificar 8 veces con la carga del divisor, lo que es fácil de conseguir con un transistor como el de la figura.

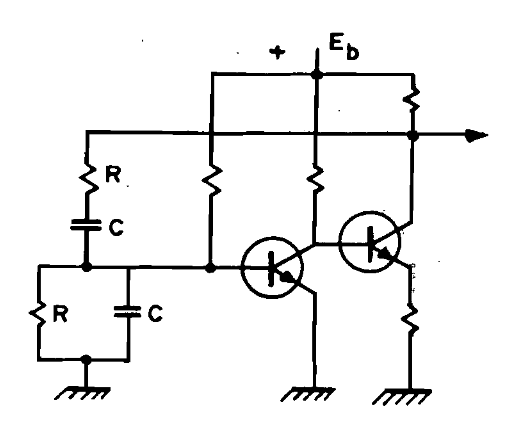


Fig. 9-6.

Otro tipo de oscilador muy popular es el de **puente de Wienn**, basado en el divisor RC-CR o puente de Wienn, representado en la realimentación del oscilador de la figura 9-6. Es fácil calcular la respuesta de frecuencia del puente de Wienn, que tiene una forma parecida a la de resonancia con un máximo de 1/3 y fase cero a la frecuencia 1/(2 π RC), que es la del oscilador. El amplificador necesita un desfase cero (por lo que es necesario usar dos etapas) y una amplificación 3, que se obtiene fácilmente.

9.5.- Oscilador Hartley.

El oscilador **Hartley** es un oscilador **LC**, con un elemento resonante en el circuito de realimentación. La originalidad del oscilador LC es la combinación de realimentación con el cambio de fase por toma intermedia en el circuito resonante (fig. 9-7).

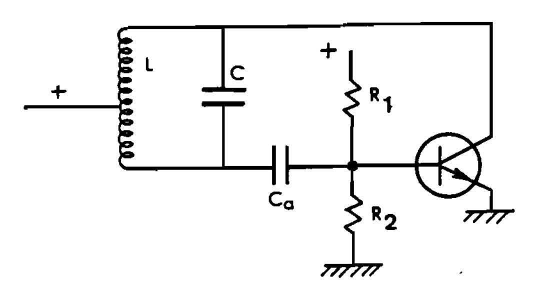


Fig. 9-7.

El amplificador es selectivo y su carga es el circuito resonante LC. La alimentación + es tierra para las señales, por lo que la fase de la fracción inferior de la autoinducción es la inversa de la superior, con lo que se consigue el signo - de β que equilibra al signo - de la amplificación, es decir, que la fase de βA es 0. La condición de amplificación 1 se cumple también con facilidad para el producto βA.

9.6.- Osciladores de cristal de cuarzo y de línea de transmisión.

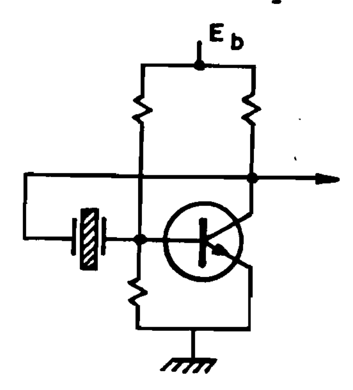


Fig. 9-8.

Cuando se desea una gran precisión en el funcionamiento de los osciladores, es conveniente usar un amplificador de banda ancha y un componente de realimentación de gran selectividad. Ese es el caso de los osciladores de cristal de cuarzo, como el de la figura 9-8. La realimentación se hace a través de un cristal de cuarzo que sólo deja que se propague su frecuencia propia, por efecto piezoeléctrico.

También puede usarse una línea de transmisión, como las que estudiaremos en un capítulo posterior, cortada exactamente a media longitud de onda de la frecuencia deseada. La señal se desfasa 180º en la línea y otros 180º en el amplificador, que compensa las pérdidas en la línea (fig. 9-9):

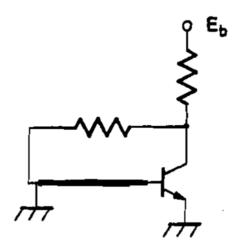


Fig. 9-9.

9.7.- Amplificadores operacionales.

Para usar bien las propiedades de la realimentación negativa, se usa un tipo de amplificador conocido como **operacional**.

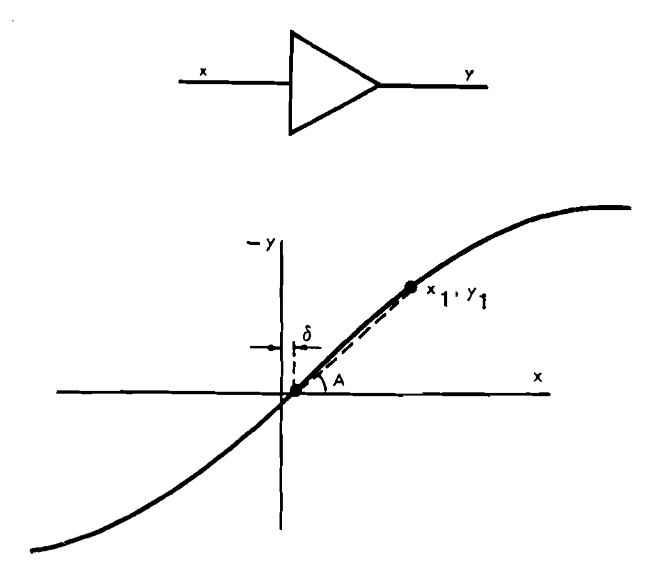


Fig. 9-10.

Un amplificador operacional es un **amplificador de c.c.** y **banda ancha** (es decir que amplifica con respuesta plana desde cero hasta relativamente altas frecuencias). La amplificación es **negativa**, es decir, es **inversor**, para incluir el signo menos necesario en el circuito de realimentación (posteriormente veremos que también se usan amplificadores diferenciales, en los que la realimentación se da siempre por la entrada negativa), y su valor absoluto debe ser muy **alto**. Se representan por un triángulo, en el que se suele hacer explícito su carácter de inversor con un signo - a la entrada. (Fig. 9-10).

La tensión de salida **y** para un valor de entrada **x** es en un momento determinado - y = A (x - δ), donde δ es la **desviación de cero**, error de entrada que tiene una gran importancia, como veremos. A es la amplificación en cada punto (ver figura 9-10). Suele ser A > 10.000 a partir del origen y hasta llegar al valor de **saturación** que representa un límite de trabajo.La saturación suele ocurrir en valores próximos a los de la alimentación. La desviación de cero suele ser inferior a un milivoltio, para conseguir lo cual es necesario recurrir a regulaciones del punto de trabajo con la temperatura, ajustes automáticos, etc...La amplificación debe mantenerse hasta frecuencias del orden de 10 Khz. o más.

Normalmente, el amplificador puede aproximarse al **amplificador operacional ideal**, con δ = 0 y amplificación y respuesta de frecuencia tendiendo a infinito.

9.8. Realimentación negativa en serie.

La realimentación negativa con amplificador operacional inversor se denomina **en serie**, distinguiéndola así de la realimentación negativa con amplificador diferencial y utilización de la entrada "directa" que veremos posteriormente.

Vamos a estudiar primero la multiplicación por un factor constante, que ,en realidad, es una **amplificación estabilizada**, que depende sólo de la realimentación, aprovechando el sencillo circuito (figura 9-11) para ver los errores producidos por las características no ideales del amplificador. Posteriormente estudiaremos circuitos más complicados con amplificadores ideales y su aplicación al cálculo analógico.

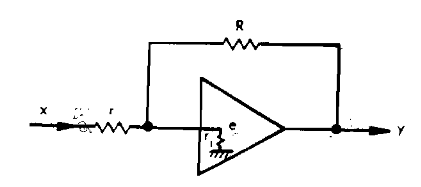


Fig. 9-11

Sean x e y las entradas y salidas del circuito, e la entrada

del amplificador (funciones del tiempo), δ y r_i la desviación de cero y resistencia de entrada del mismo. Sea A la amplificación, función de x pero acotada. Suponemos despreciable la de salida, que suele serlo frente a las resistencias de realimentación. Las ecuaciones del circuito de la figura 9-11 son:

$$\frac{x-e}{r}+\frac{y-e}{R}-\frac{e}{r_i}=0$$
$$y=-A(e-\delta)$$

Eliminando e entre las dos ecuaciones se obtiene:

$$y=-\frac{R}{r}x\frac{1}{1+\frac{1}{A}+\frac{R}{rA}+\frac{R}{r_iA}}+\delta\frac{1+\frac{R}{r}+\frac{R}{r_i}}{1+\frac{1}{A}+\frac{R}{rA}+\frac{R}{r_iA}}$$

Si se hace A muy grande, pueden despreciarse los términos en que figura A en el denominador, quedando :

$$y=-\frac{R}{r}x+\delta\left[1+\frac{R}{r}+\frac{R}{r_i}\right]$$

Si r_i es grande respecto a R y R/r es moderadamente grande, el error más importante es $\delta R/r$ referido a la salida o simplemente δ referido a la entrada. Este error es importante si la entrada es pequeña, es decir si δ no es despreciable respecto a x.

Existen algunos dispositivos especiales para disminuir ese error, casi todos ellos basados en convertir la entrada x en una corriente alterna con algún dispositivo de conmutación (ya que la c.a. no tiene desviación de cero), o bien, también por conmutación, hacer periódicamente x=0, detectar el valor de delta y compensarlo en paralelo.

9.9.- Cálculo analógico. Resolución de ecuaciones.

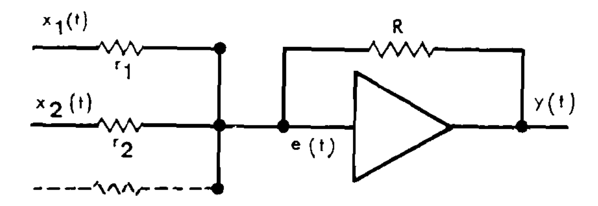


Fig. 9-12.

Vamos a describir ahora el funcionamiento del circuito de la figura 9-12. Sean x_1, x_2, ... las entradas del circuito e y la salida, todas ellas funciones del tiempo.

Escribamos la ecuación del nudo de entrada:

$$\frac{x_1-e}{r_1}+\frac{x_2-e}{r_2}+\ldots-\frac{e}{r_i}+\frac{y-e}{R}=0$$

Además es y = - Ae, y, teniendo en cuenta que A → ∞, es decir, e → 0, se verifica:

$$y=-\frac{R}{r_1}x_1-\frac{R}{r_2}x_2-\ldots$$

lo que significa que **y es una combinación lineal de las entradas.** multiplicadas por coeficientes constantes que sólo dependen de componentes pasivas, y cambiado de signo. El cambio de signo puede deshacerse fácilmente con el **inversor** de la figura 9-13, que no es más que un caso particular del circuito anterior.

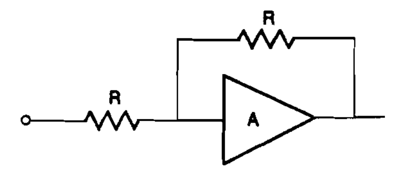


Fig.9-13.

Una aplicación inmediata del circuito anterior, primer ejemplo de **cálculo analógico**, es la resolución de sistemas de ecuaciones lineales, que se hace cableando una **analogía**, que es un circuito que responde a las ecuaciones y en el que se pueden medir directamente las soluciones. Para ello hay un procedimiento sistemático, que es como sigue:

Supongamos que queremos resolver el sistema :

$$ax_1+bx_2+cx_3=d$$
$$ex_1+fx_2+gx_3=h$$
$$ix_1+jx_2+kx_3=l$$

cableando una **analogía eléctrica** que responda a las mismas ecuaciones. El procedimiento a seguir es despejar una incógnita de cada ecuación, con lo que el sistema queda en la forma:

$$x_1=\frac{d}{a}-\frac{b}{a}x_2-\frac{c}{a}x_3$$
$$x_2=\frac{h}{f}-\frac{e}{f}x_1-\frac{g}{f}x_3$$
$$x_3=\frac{l}{k}-\frac{i}{k}x_1-\frac{j}{k}x_2$$

La figura 9-14 representa el cableado de la primera ecuación del sistema. La resistencia de realimentación R se elige arbitrariamente de un valor alto, por ejemplo 1 megohmio, de forma que la resistencia de salida del amplificador sea despreciable frente a ella. A partir de ese valor, las resistencias conectadas en serie con cada entrada se eligen de valor **inverso** al de los coeficientes en megohmios. Así por ejemplo, la primera entrada x_1 se conecta a través de r=a/b M., de forma que R/r sea entonces b/a, como el coeficiente de la ecuación. La realimentación aporta también el signo menos. La ecuación del circuito de la figura 9-14 es la primera de las ecuaciones. La entrada -d es una tensión en voltios igual al término independiente de la ecuación.

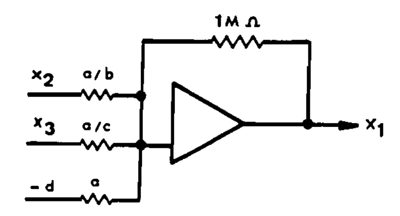


Fig. 9-14.

Del mismo modo se cablean las demás ecuaciones y se llevan (realimentan) los valores de las salidas de cada una a las entradas de las otras. La figura 9-15 es el cableado completo:

Los términos independientes, que en estas ecuaciones se suponen constantes, pueden ser funciones del tiempo cualesquiera, en cuyo caso el sistema se convierte en una transformación del grupo de variables d, h, l al x_1, x_2, x_3. Si algún coeficiente es negativo, pueden introducirse inversores como el de la figura 9-13 para hacer los cambios de signo necesarios.

9.10.- Integración y diferenciación. Ecuaciones diferenciales.

La figura 9-16 es un **integrador analógico**. Efectivamente, las ecuaciones del circuito en el plano del tiempo son:

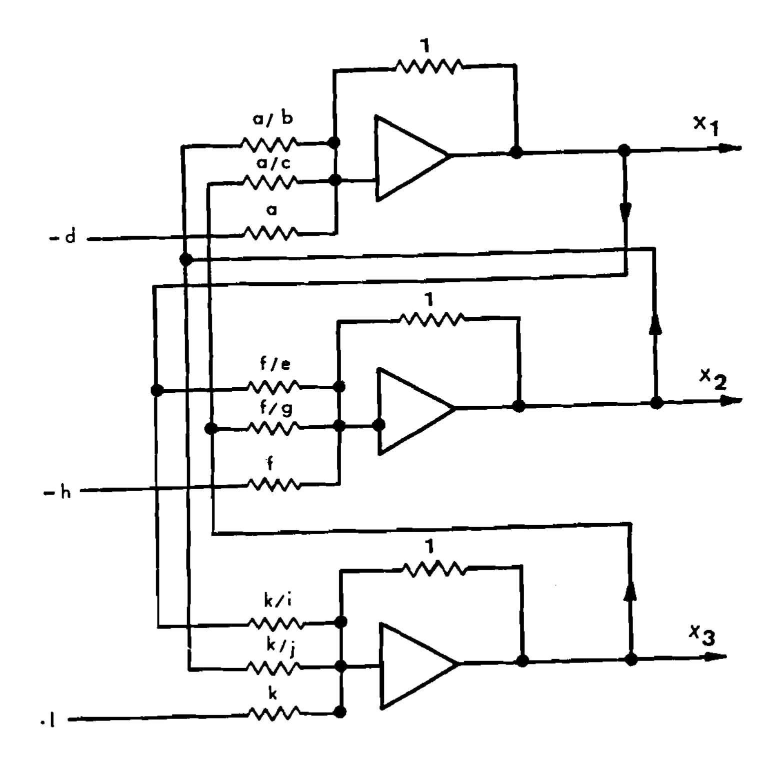


Fig. 9-15.

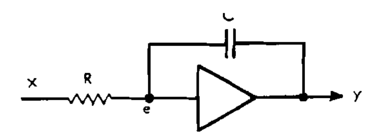


Fig. 9-16.

$$\frac{x-e}{R}+C\frac{d(y-e)}{dt}-\frac{e}{r_i}=0$$
$$y=-Ae$$

Haciendo A → ∞, es decir, e → 0, queda:

$$x=-RC\frac{dy}{dt}$$

Esta expresión indica que la **salida es proporcional a la integral de la entrada**. Si se integra , a partir de x_o, y_o, t_o, se

obtiene:

$$y=-\frac{1}{RC}\int_{t_o}^{t} x dt+y_o$$

El valor de y_0 se introduce como **carga inicial** del condensador, mediante relés conmutadores de **inicialización**. La constante RC puede hacerse RC = 1, por ejemplo, tomando R = 1MΩ y C = 1 μF.

La figura 9-17 es un diferenciador.

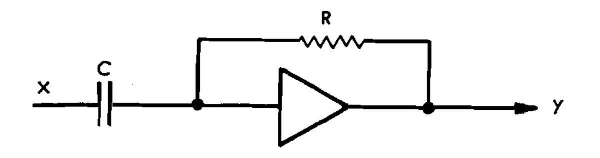


Fig. 9-17.

Si tomamos directamente e → 0, se obtiene:

$$C\frac{dx}{dt}+\frac{y}{R}=0$$
$$y=-RC\frac{dx}{dt}$$

El circuito diferenciador tiene algunos inconvenientes: A una entrada discontinua corresponde una salida infinita. Además, como la amplificación es proporcional a la frecuencia, los ruidos de alta frecuencia se amplifican demasiado. Por todo ello, en los calculadores analógicos se prefiere integrar a diferenciar.

Supongamos ahora que queremos resolver la ecuación diferencial:

$$a_o y'''+a_1 y''+a_2 y'+a_3 y=f(t)$$

donde f(t) es una función del tiempo de la que disponemos en forma eléctrica. Por ejemplo, si es una constante, una batería, si es sinusoidal, un oscilador, etc... Además conocemos los valores iniciales de y''_0, y'_0 e y_0 en el instante t=0.

Para resolver la ecuación en forma integral, despejamos la derivada de mayor orden y''', con lo que la ecuación queda:

$$y''' = -\frac{a_1}{a_o}y'' - \frac{a_2}{a_o}y' - \frac{a_3}{a_o}y + \frac{f(t)}{a_o}$$

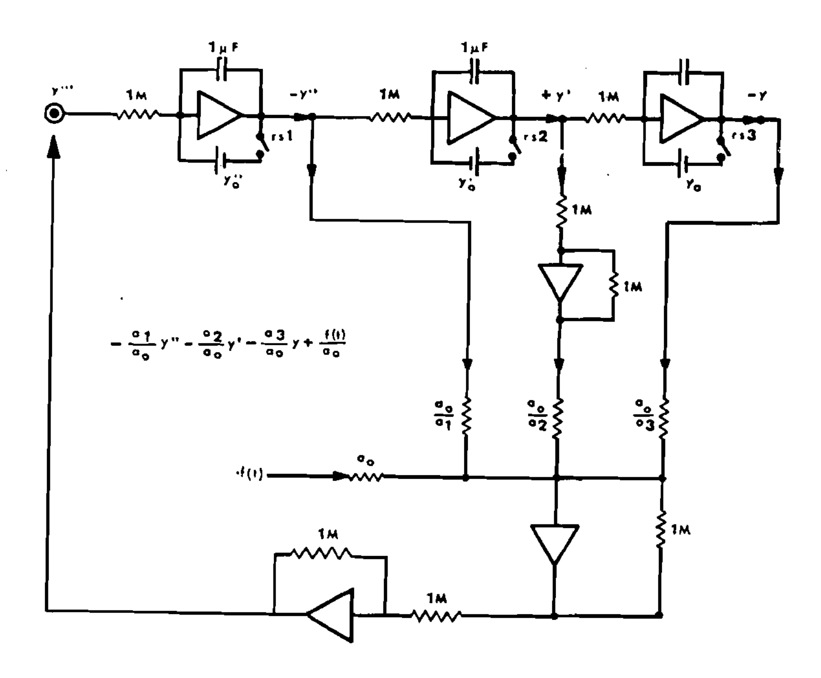


Fig. 9-18.

Supongamos que tenemos el valor y''' en la parte superior izquierda de la figura. Con tres integraciones sucesivas, obtenemos -y'', y' e y, contando con los valores iniciales que cargamos en los condensadores mediante relés accionados antes del momento t=0 y liberados en t=0.

A continuación se cablea la ecuación igual que la de la figura 9-15, haciendo la combinación lineal correspondiente e invirtiendo las variables cuando sea necesario. Obsérvese que este cableado **obtiene** el valor y''', que se realimenta al punto inicialmente considerado. El valor de f(t) se introduce en el punto adecuado.

9.11.- Multiplicación de variables.

Las operaciones no lineales son mucho más difíciles de realizar que las anteriormente descritas.

Un tipo de multiplicador bastante extendido utiliza un generador de impulsos rectangulares cuyo ancho es proporcional a una de las variables y cuya altura es proporcional a otra variable. El área de estos impulsos resultará proporcional al producto de las dos variables.

Otro procedimiento es la utilización de diodos o dispositivos de característica aproximadamente parabólica, es decir, que efectúan el cuadrado de una variable. El producto se deduce entonces (sabiendo realizar los cuadrados) por la fórmula:

$$xy=\frac{(x+y)^2-x^2-y^2}{2}$$

En general, en los calculadores analógicos se procura evitar las operaciones no lineales.

9.12.- Realimentación en paralelo.

Modernamente se tiende a reemplazar (sobre todo en operaciones individuales) el sistema de realimentación descrito por otro basado en la utilización del **amplificador operacional diferencial**, representado en la figura 9-19.

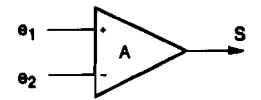


Fig. 9-19.

Este amplificador tiene la ventaja de **separar** la realimentación por un lado (su entrada -) y la entrada por otro (su entrada +), con la ventaja adicional, como veremos, de **no invertir** las señales, al pasar éstas por la entrada positiva.

Su ecuación de definición es $s = A(e_1 - e_2)$, que, cuando $A \to \infty$, obliga a que $e_1 = e_2$. para satisfacer la ecuación, a través de la realimentación negativa, como antes. Es decir, puede decirse que el amplificador da una salida s tal que, a través de la realimentación el valor de e_2 anule la entrada. La realimentación sigue siendo negativa, lo que es obligado para la estabilidad.

Consideremos en general el circuito de la figura 9-20, con un divisor de tensión de entrada y otro de realimentación:

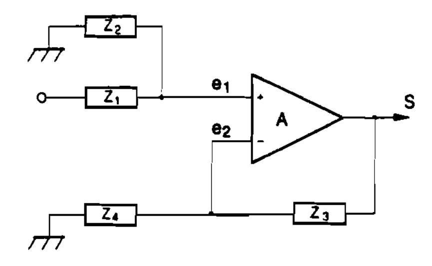


Fig. 9-20

La ecuación del amplificador se simplifica a **$e_1 = e_2$**, por ser la amplificación muy grande, como hemos visto, por lo cual, podemos escribir:

$$e_1 = e_2 = s\frac{Z_4}{Z_3+Z_4}$$
$$o\ sea$$
$$s = \frac{Z_3+Z_4}{Z_4}e_1$$

en que vuelve a aparecer que el amplificador efectúa la manipulación **inversa** a la del divisor de tensión de realimentación sin cambio de signo. Es decir, que , si hubiésemos llamado β a la realimentación ($e_2 = \beta$ s), la amplificación **desde la entrada e_1 sería $1/\beta$.**

En la misma figura hemos introducido un divisor de tensión **previo**, que actúa completamente por separado, con la ventaja adicional de la alta impedancia de entrada del amplificador operacional ideal. Incluyendo este divisor la amplificación **desde e** sería ahora:

$$s = e_1\frac{Z_3+Z_4}{Z_4} = e\frac{Z_2}{Z_1+Z_2}\frac{Z_3+Z_4}{Z_4}$$

En particular, pueden utilizarse circuitos complicados para cualquiera de las impedancias, pero es preciso tener cuidado con las de realimentación, para evitar que, al invertirse el signo para alguna frecuencia, pueda hacerse inestable.

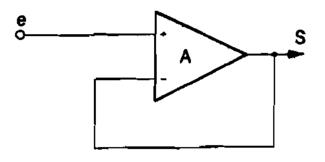


Fig. 9-21

Un ejemplo sencillo de realimentación unidad es el **seguidor** representado a la izquierda de la figura 9-21, cuya ecuación **s=e**, justifica la denominación de seguidor (en realidad es un adaptador de impedancia, ya que la impedancia de entrada es muy alta y la de salida muy baja). A la derecha de la misma figura se ve el seguidor utilizado como salida de un filtro resonante de aplicación frecuente.

CAPITULO 10

AMPLIFICADORES NO LINEALES

10.1.- Introducción.

En capítulos anteriores hemos estudiado el comportamiento **linealizado** de los amplificadores. Para ello hemos supuesto que las señales eran pequeñas desviaciones del punto de trabajo o, al menos, no se salían de la zona lineal.

En este capítulo vamos a tratar algunos problemas de no linealidad causados por **señales grandes**. Comenzaremos por hacer un estudio general, incluyendo los fenómenos de saturación y la no linealidad producida dentro de zona de trabajo. Sólo excluiremos el uso de amplificadores como conmutadores, que estudiaremos en otro lugar.

10.2.- Comportamiento de los transistores con grandes señales.

El comportamiento del transistor en grandes zonas de funcionamiento se llama comportamiento **dinámico**, aunque existe alguna confusión porque también se usa esa denominación para respuesta a transitorios. Sin embargo, conservaremos esa denominación por su carácter tradicional, aclarando que nos referimos a grandes señales.

El comportamiento del transistor está definido por sus características de colector, como se ve en las figuras 10-1 y 10-2. En ambas se ha tomado como abscisa el valor V_C y como ordenada el de I_C. En la figura 10-1 se ha tomado V_{BE} como parámetro, mientras que en la 10-2 el parámetro es I_B.

Según las fórmulas teóricas, la corriente de colector no depende de V_C, por lo que ambas figuras deberían mostrar curvas horizontales . La pendiente que aparece es debida a la existencia

de una resistencia finita de colector, de valor $1/h_{22}$ .La tensión de colector tiene una pequeña influencia en la difusión de cargas a través de la base, y este efecto se traduce en la pendiente de las curvas.

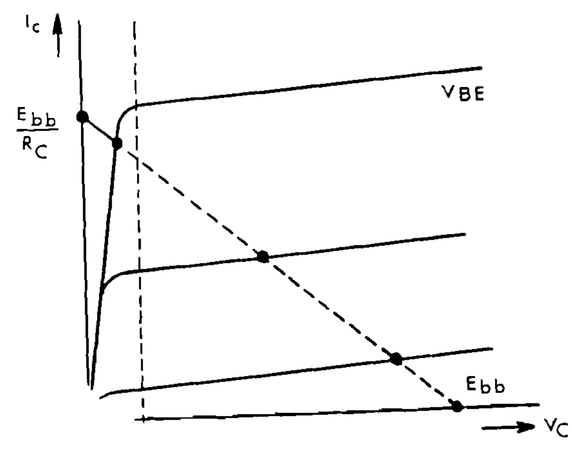


Fig. 10-1.

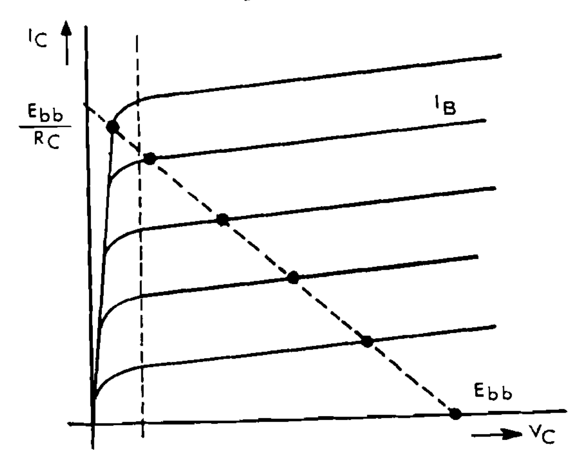


Fig. 10-2.

Por lo demás, las curvas de la figura 10-1 tienen una distribución exponencial, mientras que las de la figura 10-2 tienen una distribución lineal, como ya vimos en un capítulo anterior. La zona de la izquierda de ambas figuras corresponde a la **saturación**. Al no haber una tensión de colector suficiente, o ser excesiva la corriente de base o la tensión de base, el transistor está **saturado**, conduce y se comporta como una pequeña resistencia, que es el valor inverso de la pendiente de la recta de saturación.

Dada una recta de carga cualquiera, $V_C = E_{bb} - R_C I_C$, la intersección de esta recta con las características da los puntos de trabajo.

Despreciando de nuevo el valor de h_{22}, podemos calcular el valor de I_C en función de V_{BE} por la ecuación :

$$I_C = Ke^{\frac{q}{KT}V_{BE}} = Ke^{39\,V_{BE}}$$

recordando que q/KT vale 39 V^{-1} a temperatura ambiente.

La figura 10-3 representa un amplificador elemental y la figura 10-4 es la representación gráfica de I_C en función de V_{BE}. Para estudiar la linealidad puede tomarse I_C como variable de salida, ya que la verdadera salida del amplificador V_C depende de I_C a través de la ecuación lineal $V_C = E_{bb} - R_C\, I_C$. El crecimiento de I_C es exponencial (rapidísimo) y extraordinariamente no lineal. El valor se triplica cada 25 milivoltios, como se indica en la figura por encima y por debajo de un "punto de trabajo", que puede ser 0,5 V., 1 mA. A partir de cierto valor (puede seguirse a lo largo de la recta de carga de la figura 10-1) se acerca rápidamente a la saturación, que es un valor límite de I_C que vale $I_C = E_{bb} / R_C$.

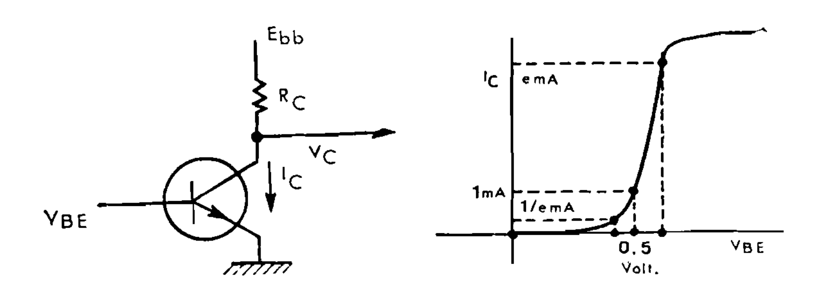


Fig. 10-3. Fig.10.4.

En la figura 10-5 se representa la salida del amplificador en función de la entrada. Llama la atención el decrecimiento tan rápido de la curva (exponencial invertido) y el poco margen que hay para elegir una zona linealizada. Prácticamente, y sobre todo si se tiene en cuenta que además hay una dependencia de la temperatura, este amplificador no es utilizable en zona lineal.

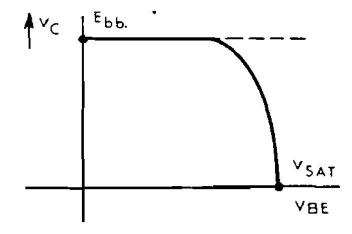


Fig. 10-5.

Muy distinto es el caso de tomar como variable de entrada I_B, ya que entonces el transistor sí se comporta linealmente en una extensa zona, por ser aproximadamente $I_C = \beta I_B$. El transistor es un **amplificador lineal de corriente**, pero se usa pocas veces como tal, además de tener el inconveniente de depender de β, distinta para cada transistor.

10.3.- Linealización y estabilización de la amplificación.

Ya se vio que la realimentación estabilizaba la amplificación, haciéndola independiente de los parámetros del transistor. Además, también linealiza su comportamiento, al depender de otras componentes lineales.

El procedimiento más extendido de linealización y estabilización de la amplificación es el uso de una **resistencia de emisor**, que, en realidad, es una forma sencilla de realimentación. Sea el circuito tipo de la figura 10-6, análogo al de la figura 10-3, pero con una resistencia de emisor **r**.

Tomaremos aproximadamente $\alpha = 1$, y despreciaremos los valores de I_{co} y h_{22} . Tomaremos $q/KT = 39\ V^{-1}$.

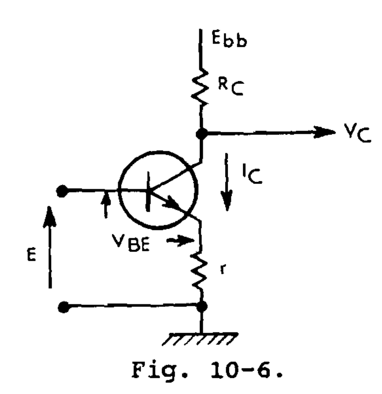


Fig. 10-6.

Las ecuaciones del circuito serán:

$$E = rI_C + V_{BE}$$
$$I_C = Ke^{39 V_{BE}}$$
$$V_C = E_{bb} - R_C I_C$$

Si tomamos por comodidad I_C como variable de salida (recordando que la dependencia de la salida V_C de I_C es totalmente lineal), despejamos E en función de I_C en la forma:

$$E = rI_C + \frac{1}{39}\ln\frac{I_C}{K}$$

Normalmente no se conoce K, sino una pareja de valores del punto de trabajo, I_o, V_o, que definen K a través de la relación:

$$I_o = Ke^{39V_o} \text{ o sea } V_o = \frac{1}{39}\ln\frac{I_o}{K}$$

Ahora podemos ya despejar la entrada en función de la salida:

$$E = rI_C + \frac{1}{39}\ln\frac{I_C}{K} = rI_C + \frac{1}{39}\ln\frac{I_C I_o}{KI_o} = rI_C + \frac{1}{39}\ln\frac{I_o}{K} + \frac{1}{39}\ln\frac{I_C}{I_o}$$

con lo que queda :

$$E = rI_C + V_o + \frac{1}{39}\ln\frac{I_C}{I_o}$$

que es la fórmula que nos permite sacar las siguientes conclusiones:

El primer término es perfectamente **lineal**, depende sólo del valor de **r.** Si se despreciasen los demás términos, la amplificación sería R_C / r, también dependiente sólo de estas dos componentes pasivas. Este es el término verdaderamente estabilizador.

El término V_o es una constante del orden de 0,5 voltios. Varía con la temperatura a razón de aproximadamente -2,5 mV / º. El último término es muy pequeño, cero si I_C = I_o y su variación es también pequeña.

Se llama **coeficiente de estabilización** al valor :

$$s = \frac{rI_o}{rI_o + V_o}$$

que es la relación de rI_o a E para el punto I_C = I_o .

Recordaremos una vez más que el inconveniente de hacer r, y por consiguiente s, grande, es que la amplificación se reduce al valor R_C / r, que es el precio que se paga por la estabilización.

La figura 10-7 es la representación de E = f(I_C). En él se toma I_C como abscisa y E como ordenada lo que simplifica algo

los cálculos, aunque viene a ser una representación de amplificación inversa. El parámetro es r. La curva inferior es el caso r=0, es decir, sin estabilización. La pendiente es enorme, la amplificación del orden de 400 (recuérdese que para visualizar la amplificación hay que pensar en la pendiente inversa, dI/ dE y multiplicar por R_C) , pero se ha señalado la enorme influencia de una variación de 20 ºC, que desplaza la curva 0,05 mV, que es la variación de la curva entre 1 y 10 mA.

Para r = 200 ohmios, la curva es más lineal y el desplazamiento señalado con la misma variación de temperatura es despreciable. El coeficiente de estabilización para esa curva es:

$$s=\frac{0,2}{0,7}\approx 0,3 \text{ para } 1 \text{ mA.}$$
y mucho mejor ,
$$s=\frac{2}{2,55}\approx 0,8 \text{ para } 10 \text{ mA.}$$

Los cambios relativos de los términos V_0 y 1/39 ln I/I_0 debidos a **cualquier causa** influyen en la proporción 1-s.

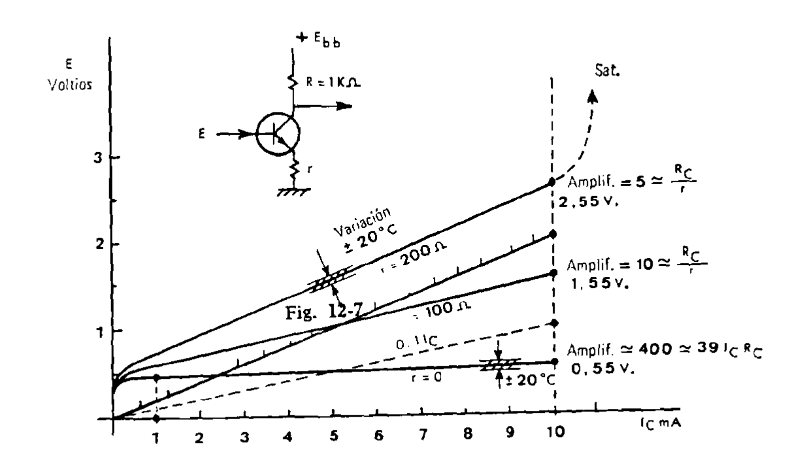


Fig. 10-7.

10.4.- Alinealidad de los amplificadores de transistores.

Utilizando las curvas de la figura 10-7 se puede calcular la salida que corresponde a cualquier entrada. Es fácil comprender que la zona utilizable es bastante extensa si se estabiliza con una resistencia de emisor y se elude la zona de saturación donde la curvatura es grande, a bajos valores de E.

La estimación cuantitativa de la **alinealidad** varía de unos autores a otros. Nosotros definiremos la alinealidad como la variación relativa de la amplificación por unidad de la variación

de entrada E. Este parámetro se calcula de la siguiente manera:

$$E=V_o+rI_C+\frac{1}{39}\ln\frac{I_C}{I_o}$$
$$\frac{dE}{dI_C}=r+\frac{1}{39}I_C$$

que puede reconocerse fácilmente como la impedancia generalizada de emisor. Ahora calcularemos la amplificación y su variación relativa:

$$A=-\frac{R_C}{r+\frac{1}{39}I_C}$$
$$A_L=\frac{1}{A}\frac{dA}{dE}=\frac{1}{A}\frac{dA}{dI_C}\frac{dI_C}{dE}$$
$$A_L=\frac{r+\frac{1}{39I_C}}{-R_C}\left[\frac{R_C}{[r+\frac{1}{39I_C}]^2}\right](-\frac{1}{39I_C^2})\frac{1}{r+\frac{1}{39I_C}}$$
$$A_L=\frac{\frac{1}{39I_C^2}}{(r+\frac{1}{39}I_C)^2}=\frac{39}{(1+39I_Cr)^2}$$

Si suponemos ahora que la estabilización es tal que $39I_Cr >>1$, podemos despreciar el 1 en el denominador y queda:

$$A_L=\frac{1}{39(I_Cr)^2}$$

Si, por ejemplo, dejamos que la caída de tensión I_C r en el punto de trabajo sea 1 voltio, A_L será 0,025, es decir, la amplificación variará un 2,5% por cada voltio que varíe la entrada.

En un amplificador normal se suele fijar la variación total a la salida del orden de 0,8 E_{bb} (dejando un margen para alejarse de la saturación y el corte). A partir de ahí, con una amplificación A, la variación total de la entrada será 0,8 E_{bb} /A y el valor de A_L da la alinealidad total en todo el recorrido de entrada. Es corriente permitir que este valor llegue a un 5%.

A veces, al parámetro A_L se le llama distorsión y está directamente relacionado con la distorsión que hemos definido anteriormente, si bien no coincide con ella.

En la figura 10-8 se representa la corriente I_C en función de la entrada E. La curva de funcionamiento "normal" (según las ecuaciones) **limita** por un extremo (E bajo) en que se anula y no puede hacerse negativa, y, por el otro lado (E alto) en la saturación, que ocurre para el valor :

$$I_C=\frac{E_{bb}}{R_C+r}$$

para el cual se anula V_C y el transistor queda simplemente cortocircuitado. Está claro que el punto de trabajo normal debe elegirse alejado de ambos extremos, en un punto como el A de la figura.

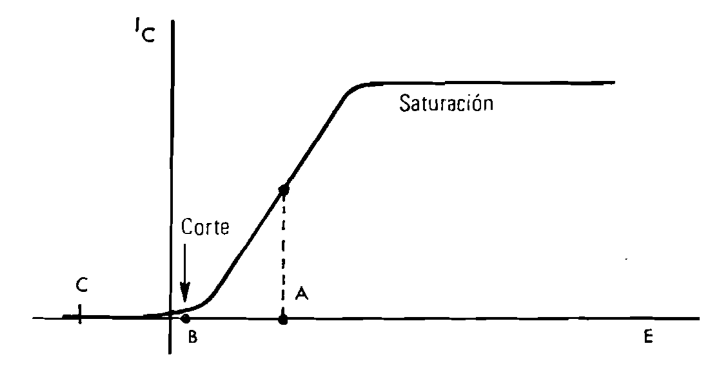


Fig. 10-8.

10.5.- Amplificadores de clase B, AB y C.

A veces se utiliza intencionadamente el corte de la figura 10-8 para limitar las señales, en instrumentación o comunicaciones. Según la polarización y utilización de este efecto, se clasifican en :

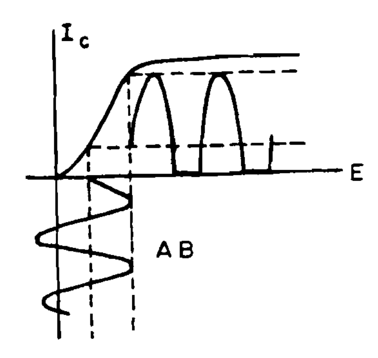


Fig. 10-9.

Clase A: Están polarizados en zona lineal evitando el corte y la saturación. Son los que hemos estudiado hasta ahora.

Clase AB: Se polarizan como en clase A, pero se usan señales que llegan al corte, produciéndose un corte de las crestas negativas, como se ve en la figura 10-9. Es un **limitador** de señales excesivas.

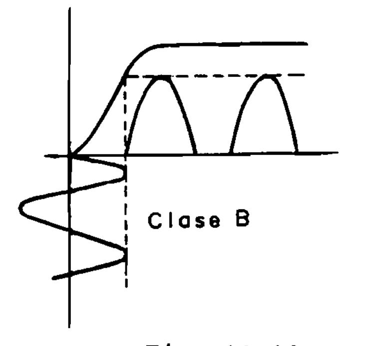


Fig. 10-10.

Clase B: Se polarizan al corte Las señales son **rectificadas**, como con un diodo, pero con amplificación. (Fig. 10-10)

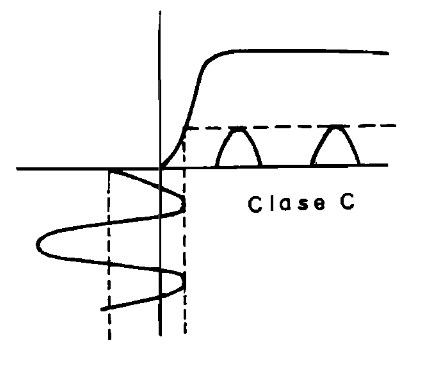


Fig. 10-11.

Clase C: Si se polariza por debajo del corte ,sólo se amplifican las crestas, como se ve en la figura 10-11. De esta forma se detectan las señales grandes, ignorando las que no llegan al corte.

Tanto el amplificador tipo B como el C no consumen en ausencia de señal. Por supuesto no son lineales, pero, debido a su distorsión, hacen aparecer gran cantidad de armónicos, que permiten generar frecuencias múltiplos de la original, lo que es útil en algunos casos, además de sus propiedades limitadoras.

CAPITULO 11

CIRCUITOS DE CONMUTACION

11.1.- Generalidades.

En este capítulo estudiaremos circuitos que utilizan componentes en condiciones límites de corte o saturación. Algunos funcionan cambiando la forma de las señales y otros (o los mismos a la vez) amplifican en el sentido de obtener más potencia de señal a la salida de la que reciben a la entrada. En este capítulo introduciremos algunos circuitos cuya función es **lógica**, si bien dejaremos para otro capítulo el estudio de los **circuitos digitales**, quizá el caso particular más importante de conmutación, pero que, por su enorme importancia actual necesita un estudio por separado. Aunque nos concentraremos en circuitos eléctricos, introduciremos el relé como antecesor de los componentes de conmutación, y todavía actual en algunas aplicaciones.

11.2.- Relés.

Un relé consiste en un electroimán que atrae una armadura magnética que acciona mecánicamente uno o varios conmutadores que aquí denominaremos "contactos". La figura 11-1 representa la disposición típica de un relé.

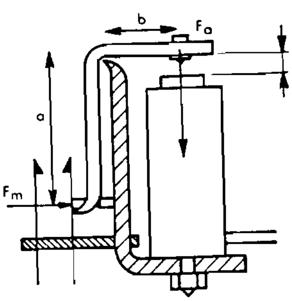


Fig. 11-1

Cuando la fuerza de atracción sobre la armadura F_a es mayor que la que opone el muelle a/b F_m la armadura se pone en movimiento. Si se cumple la condición adicional:

$$\frac{dF_a}{dx} > \frac{a}{b}\frac{dF_m}{dx}$$

el movimiento ya no cesa hasta que la armadura llega al final de su recorrido, cambiando los contactos correspondientes. Esta condición garantiza que no haya posiciones intermedias y el relé se comporte de forma **biestable**, sólo pueda quedarse en posiciones extremas. Si se utiliza el relé para conmutar una tensión E_b determinada y representamos en la figura 11-2 la salida (en el contacto) en función de la entrada (intensidad en la bobina), se obtiene para el relé cuya representación es la de la izquierda de la figura la curva de **histéresis** de la derecha. Al llegar al valor I_2, el relé se **acciona**, pero no se suelta hasta bajar a un valor menor, I_2.

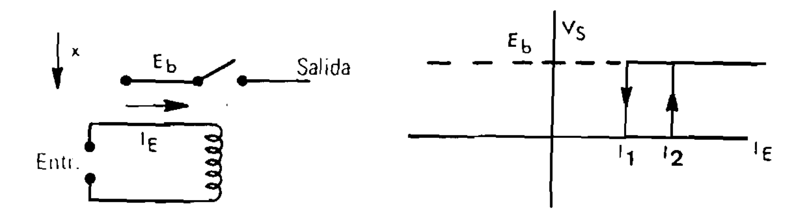


Fig. 11-2

El relé tiene varias aplicaciones características, que lo hacen insustituible en algunos casos:

a) **Separa galvánicamente** los circuitos de salida y entrada. No hay conexión entre ambos. Por ejemplo, la salida puede ser una corriente alterna y la entra c.c. El nivel o polarización de entrada es totalmente independiente de la salida.

b) Amplifica en potencia, ya que la potencia conmutada a la salida puede ser mucho mayor que la aplicada a la entrada. Los relés de aplicación electrotécnica de gran potencia se llaman **contactores**.

c) Debido al ciclo de histéresis, se puede polarizar el relé con una intensidad intermedia entre I_1 e I_2 (fig. 11-2), con lo que la posición del relé queda indeterminada. Si se sube la intensidad momentáneamente sobre I_2, se queda accionado y si se baja momentáneamente por debajo de I_1, se quedará suelto. Esto significa que el relé así polarizado tiene **memoria**. Puede "escribirse" la posición de accionado o suelto, tal y como hemos dicho, y "leerla" (observando la posición) o cambiarla. Esto se denomina una **función lógica de memoria**, y ha sido el precedente de otros circuitos de memoria que estudiaremos más adelante.

11.3.- El transistor como conmutador.

El transistor es un conmutador controlado si se usa **sólo** en el **corte** (tensión de base muy baja o cero) y en **saturación** (con tensión o intensidad suficientemente altas), prescindiendo de la zona de trabajo usada hasta ahora. El transistor no es un conmutador biestable por naturaleza, ya que, si no se cuida la entrada, puede entrar en zona de trabajo. El consumo de potencia (producto de la intensidad de colector o emisor, aproximadamente iguales, por la tensión de colector) es muy bajo en conmutación por ser prácticamente cero la tensión o la intensidad. Esto permite usar transistores integrados de tipo bipolar o MOSFET muy pequeños para conmutar potencias mucho mayores que las permisibles en zona lineal.

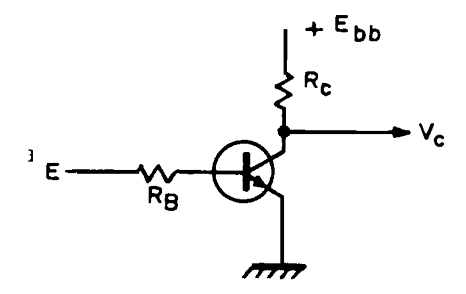


Fig. 11-3.

El circuito conmutador básico de transistor bipolar es el de la figura 11-3, llamado también **inversor**, ya que la salida es baja cuando la entrada es alta y viceversa. Si a la entrada se da un **cero lógico** o bajo (**low**), que debe estar por debajo de la tensión de base que permite una corriente apreciable de colector, el transistor está cortado o abierto, **no conduce**, y la salida de colector es la tensión de alimentación E_{bb}, es decir, es un **uno lógico**, o alta (**high**), con una impedancia de salida R_C.

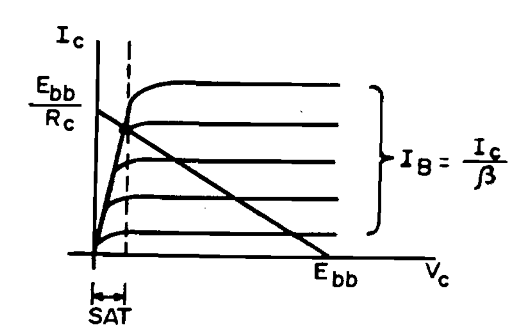


Fig. 11-4.

En las características de la figura 11-4, el punto de salida correspondiente al **corte** es el punto $E_{bb}, 0$. Puede obtenerse con $I_B = 0$ o con $V_{BE} = 0$. El valor de R_B no es crítico para ese punto. En general, se desprecia I_{co} y también se supone $I_E = I_C$.

Esta situación de corte es uno de los estados del conmutador, con entrada baja y salida alta. La impedancia de salida es R_C, que debe ser suficientemente baja.

Para **saturar** el transistor, es necesario dar un valor de I_B, a través de R_B suficientemente alto para que el transistor esté en zona de saturación, que es la zona de la figura 11-4 señalada a la izquierda, donde las curvas características son ascendentes hasta llegar a su valor de "zona de trabajo", que es βI_B. Eso quiere decir que en esa zona es $I_C < \beta I_B$, o, lo que es lo mismo, $I_B > I_C / \beta$. El punto de saturación es aproximadamente el corte de la línea de carga con el eje de ordenadas, es decir, el punto 0, E_{bb} / R_C. La **condición de conmutación** se convierte ahora en :

$$I_B > \frac{I_C}{\beta} \; y \; como \; I_C = \frac{E_{bb}}{R_C} \; y \; I_B \approx \frac{E}{R_B} \; ya \; que \; V_{BE} \approx 0$$

$$\frac{E}{R_B} > \frac{E_{bb}}{\beta R_C} \; y como E_{bb} > E$$

$$R_B < \beta R_C$$

Normalmente se proyecta tomando una cota para $\beta >> 10$ y se toma $R_B = 10\ R_C$, con lo que se cumple seguro la condición de conmutación.

En estado de saturación, el transistor se comporta como un cortocircuito. La impedancia de entrada sigue siendo R_B, como en el corte, pero la impedancia de salida es mucho menor,del orden de un ohmio, propiedad que se utiliza en algunas aplicaciones.

El MOSFET puede también utilizarse como conmutador, en forma análoga al transistor, pero con algunas ventajas adicionales. La figura 11-5 es un inversor con MOSFET de enriquecimiento de canal n (el más usado), junto con sus curvas características.

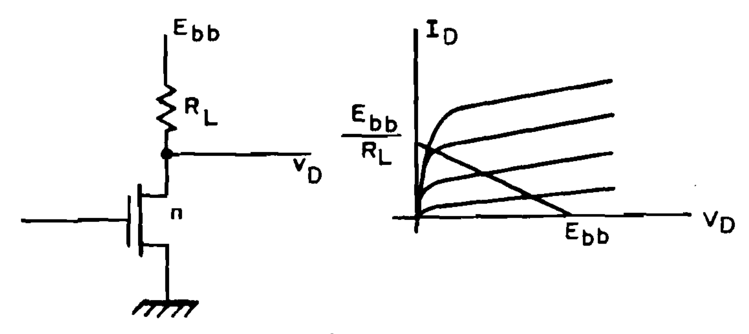


Fig. 11-5.

Esta vez **no** es necesaria la resistencia de entrada y la corriente de puerta es prácticamente nula, aunque se apliquen valores altos de tensión de entrada. Si la entrada es **baja**, **(low)** equivalente a un **0** digital, el MOSFET no conduce y la salida es

alta, (**high**), igual a E_{bb}, o a un **1** digital. La impedancia de salida es R_L. Si la entrada es alta (puede ser tan alta como E_{bb}), el MOSFET conduce, la salida es baja (moderadamente, la saturación es algo menos eficaz que en el transistor bipolar), del orden de un voltio, la impedancia de salida también es moderadamente baja, de unos ohmios. Pero el único inconveniente del MOSFET como conmutador es su menor velocidad de conmutación, debido a la inevitable capacidad de puerta, que lo hace uno o dos órdenes de magnitud más lento que el transistor bipolar.

La figura 11-6 representa dos variantes del inversor MOSFET. En la primera, la conmutación se hace con MOSFET complementarios, **CMOS**. En el circuito de la derecha, la carga se ha sustituido por otro transistor MOSFET, cuya conducción se garantiza polarizando la puerta al polo positivo. En ambos casos se evita el uso de resistencias, lo que da ciertas ventajas durante el proceso de fabricación.

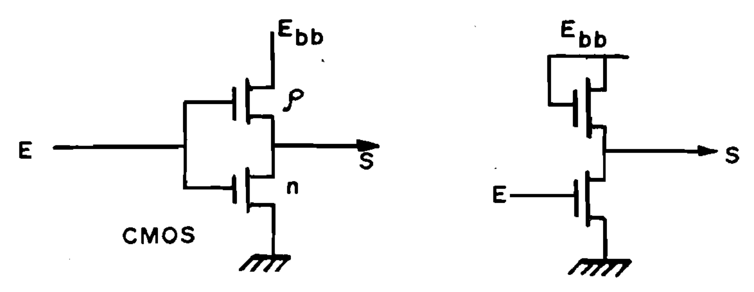


Fig. 11-6.

Por último, el MOSFET permite la **conmutación en serie**, que aunque también era posible en transistores bipolares, se hacía mucho más complicada. El esquema de la figura 11-7 presenta un conmutador de paso con MOSFET.

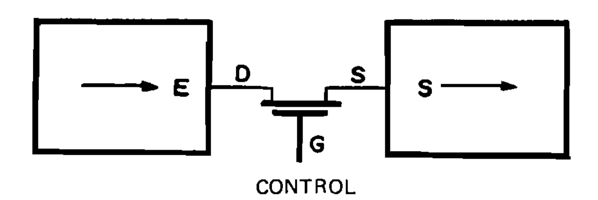


Fig. 11-7.

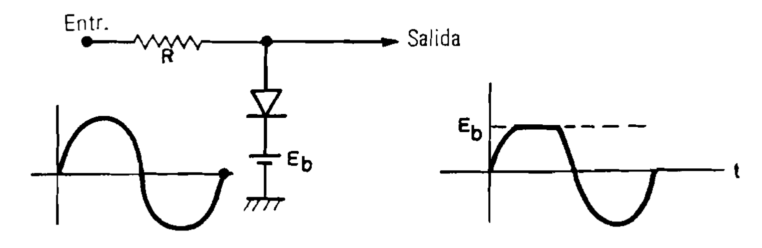
Si la tensión de puerta es alta, el conmutador conduce, y la salida se hace igual a la entrada. Si la puerta es baja, el conmutador no conduce, y la salida es cualquiera, independiente de la entrada.

Colocando un conmutador- serie a la salida de otros dispositivos, puede obtenerse también un **tercer estado**, en que la salida es independiente del circuito anterior.Se llega así a la **lógica de tres estados, TSL** (three states logic). La salida puede ser alta, baja o independiente del circuito conmutador.

11-4.- Cortadores y restablecedores.

En este apartado vamos a estudiar algunos **formadores de onda**, que utilizan conmutadores, si bien no en forma digital.

El circuito de la figura 11-8 es un **cortador**, que elimina la parte de la tensión de entrada que supera a la tensión E_b de la batería de referencia. Cuando la tensión de entrada es mayor que E_b, el diodo conduce y se porta como un cortocircuito, dando la tensión de la batería a la salida. Cuando la entrada es inferior a E_b, el diodo no conduce y la salida es igual a la entrada. Del mismo modo, la figura 11-9, es un **cortador**, que deja pasar la parte de la entrada que es superior a E_b, lo contrario de la anterior. La figura 11-10 es una combinación de ambas. Todos estos circuitos actúan como **limitadores** de señal.



Fiq. 11-8.

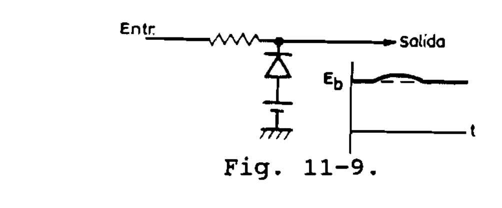


Fig. 11-9.

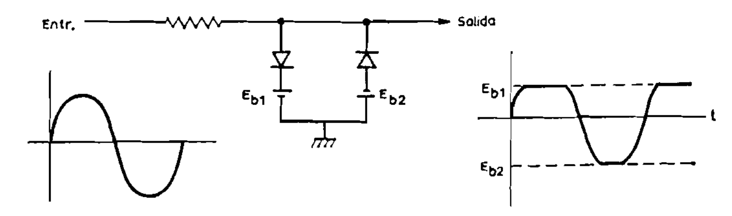


Fig. 11-10.

En un capítulo anterior hemos estudiado cómo se cambia el nivel de polarización o continua de una onda, con un CR. La figura 11-11 es un circuito de este tipo, llamado **restablecedor.**

La figura 11-12 es un **restablecedor automático**. El diodo hace que el condensador se cargue muy rápidamente a la tensión más negativa que aparezca a la entrada y, a partir de ese momento, le suma esa tensión, con lo que la salida se hace siempre positiva. El obtener una señal siempre positiva es interesante, por ejemplo, si se quiere amplificar luego en un

circuito que sólo admita señales positivas. La resistencia indicada con trazos en la figura hace que el condensador se descargue al cabo de un tiempo grande respecto a la señal, pero que permita iniciar de nuevo el proceso con otras señales.

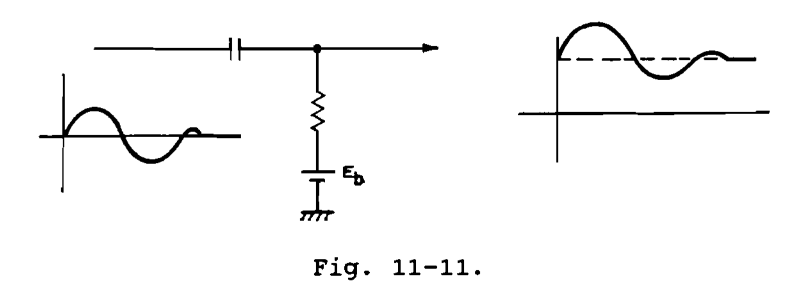


Fig. 11-11.

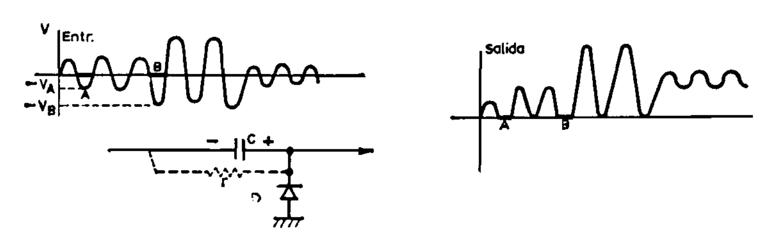


Fig. 11-12.

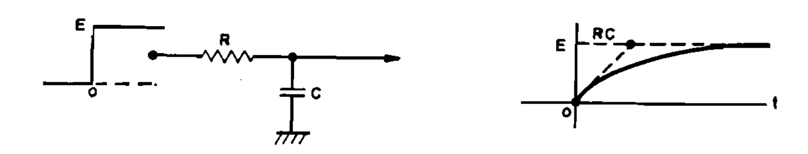


Fig. 11-13.

11.5.- Generadores de dientes de sierra.

La carga de un condensador a través de una resistencia es un fenómeno ampliamente conocido y responde a una ecuación diferencial de primer orden, cuya solución está representada en la figura 11-14.

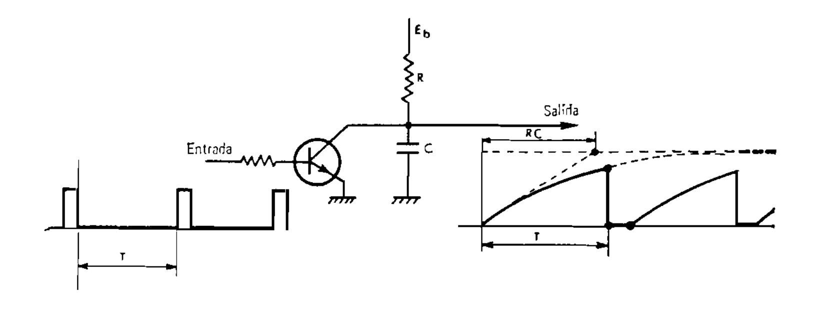


Fig. 11-14.

La solución de la ecuación es :

$$f(t)=E(1-e^{-\frac{t}{RC}})$$

Esta curva es una exponencial invertida, la tangente en el origen corta a la asíntota en el valor T = RC (constante de tiempo). La pendiente de la curva en el origen es pues E/RC. La curva se aleja (lentamente al principio) de la tangente, de forma que, al cabo del tiempo T, la tangente alcanza el valor E y la curva el valor E(1 - 1/e) = 0,63 E (casi un 40%). Sin embargo, al cabo del tiempo 0,1 T, la tangente toma el valor 0,1 E y la curva $E(1 - e^{-0,1})$ = 0,095 E, es decir, difieren sólo en 0,005E, o sea un 5%.

Un **generador de dientes de sierra** es un circuito como el de la figura 11-14. Los impulsos de entrada cortocircuitan el transistor dando salida cero. Cuando la entrada se hace baja, el condensador se carga a través de la resistencia R, con una curva de forma exponencial, como la dibujada. Si se desea que la salida sea muy lineal, debe usarse sólo una pequeña fracción de la constante de tiempo RC, con lo que, como hemos visto antes, la curva se aproxima mucho a la tangente. El circuito es un generador de señal triangular (en diente de sierra) sincronizada con la entrada.

El generador de diente de sierra se utiliza para los **circuitos de barrido** en instrumentación (osciloscopios), radar, televisión, etc...

Existen varios procedimientos para mejorar la linealidad del diente de sierra (en aplicaciones donde sea crítico que la subida sea recta). Los más conocidos son :

a) **Integrador analógico.** El circuito de la figura 11-15 es un integrador analógico inicializado de forma análoga al sincronismo de la figura 11-14. El condensador se descarga con el sincronismo y, cuando se libera, integra una entrada constante dando una salida en rampa invertida respecto a la señal. Antiguamente se llamaba a este circuito "circuito Miller".

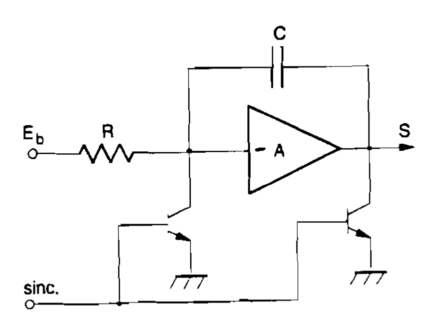


Fig. 11-15.

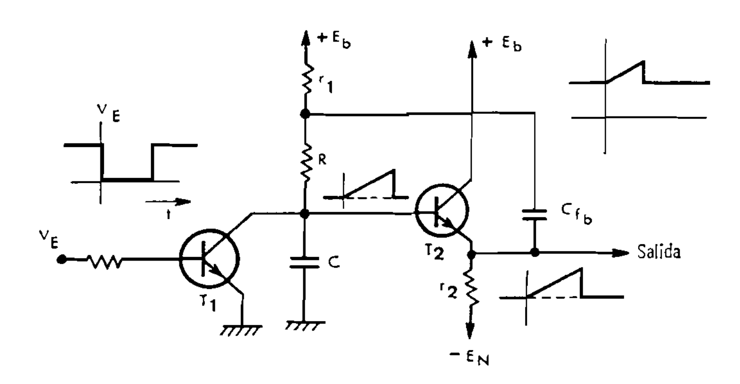


Fig. 11-16.

b) **Circuito Bootstrap.** Este circuito realimenta la salida para conseguir la carga del condensador a intensidad constante. Generalmente se hace con amplificación unidad (ver figura 11-16), con un seguidor de emisor. El transistor T_1 es un sincronizador. Cuando comienza la carga del condensador C a través de R (r_1 se considera despreciable frente a R), el transistor T_2 sigue la carga y, a través del gran condensador C_{fb} aumenta la tensión de alimentación en la confluencia de r_1 y R, manteniendo la caída de tensión en R aproximadamente constante, con lo que la intensidad de carga también lo será.

11-6.- Multivibradores.

Los **multivibradores** son circuitos que sólo pueden tomar estados extremos y en que esta característica de conmutación está conseguida a través de una fuerte realimentación positiva que hace **inestables** los estados intermedios. Con ello, el transistor, que es un excelente conmutador pero que no es biestable por naturaleza (se le "utiliza" como conmutador) se hace biestable, aunque es necesario el uso de dos transistores para conseguirlo.

Existen las siguientes clases de multivibradores, cuya funciones son diferentes, y que describiremos en párrafos posteriores con mayor detalle.

a) Multivibrador estrictamente **biestable**, que permanece en uno de los dos estados biestables que le son propios hasta que se le haga cambiar desde el exterior. Es el multivibrador **biestable** propiamente dicho, llamado también **flip-flop**, o, algunas veces, circuito báscula. Su característica es que tiene **memoria**.

b) Multivibrador **monoestable**. Está en reposo en un estado determinado. Cuando recibe un impulso adecuado cambia de estado y permanece en el otro estado un tiempo determinado. Es un **temporizador**.

c) Multivibrador **libre** o **astable**, que cambia rítmicamente de un estado al otro. Es un **reloj**.

d) Circuito Schmitt o **Schmitt-trigger**, que toma un estado u otro según el valor de la tensión de entrada, convirtiendo así una función cualquiera en función discreta (con sólo dos valores posibles) del tiempo. Sirve para fijar **umbrales** de conmutación.

11-7.- Circuito flip-flop.

El circuito flip- flop es un circuito biestable. En cada una de los dos estados extremos uno de los transistores (figura 11-17) conduce y el otro no.

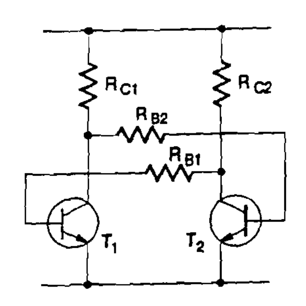


Fig. 11-17.

Si el transistor T1 de la figura 11-17 conduce, está cortocircuitado y da tensión baja a la base del transistor T2, que permanece cortado. A su vez, la tensión en el colector de T2 es alta (recuérdese que las resistencias de colector son pequeñas respecto a las de base) y, a través de la resistencia de base de T1, mantiene a T1 conduciendo. Naturalmente, debe cumplirse la condición de conmutación, es decir, que las resistencias de base sean de unas 10 veces las de colector.

El estado opuesto, T1 no conduciendo y T2 conduciendo, es igualmente estable. Todos los estados intermedios son inestables porque los dos inversores conectados entre sí dan una fuerte realimentación positiva.

Los dos **estados** pueden denominarse **0 y 1**. Por ejemplo, si T1 conduciendo es 0 y T2 conduciendo es 1, podrían sustituirse las resistencias de colector por dos bombillas que iluminasen un 0 y un 1. O sólo la de T2 con un 1 y que estaría apagada en el estado 0.

Aplicando un corto impulso positivo en la base de T1 se le obliga a conducir y a cortar a T2, provocando el estado 0. Podemos denominar a esa entrada E_1 **"escribir 0"** o **"borrar"**. La entrada E_2 provoca el estado 1 y puede denominarse **"escribir 1"** o **activar** el flip-flop.

Lo que haya escrito (estado 0 o 1) permanece estable hasta que se cambie desde el exterior. Midiendo la tensión en el colector de T1 se tiene un valor alto (T1 no conduce) en el

estado 1. El colector de T1 es la **salida F** del flip-flop, bajo en el estado 0 y alto en el estado 1. El colector de T2 está alto cuando el estado es 0 y bajo cuando es 1. Se le llama **salida no F**, es la inversa digital de la salida F.

Por último, si se da un impulso suficientemente corto en ambas bases a la vez se actúa sólo sobre el transistor que estuviese en ese momento cortado, haciéndolo conducir. Con este tipo de entrada se **cambia** el estado, de 0 a 1 o de 1 a 0, según lo que hubiese antes.

11.8.- Multivibrador monoestable.

La figura 11-18 representa un multivibrador monoestable, que permanece normalmente en uno de sus estados extremos.

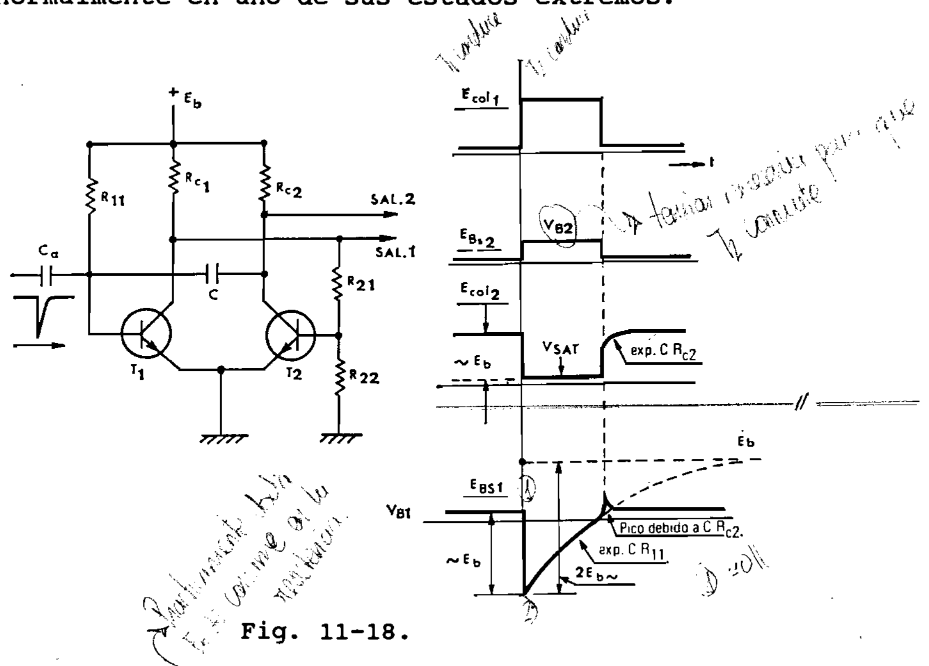


Fig. 11-18.

Normalmente, el transistor T1 conduce, polarizado a través de la resistencia R_{b1}, al valor V_{B1}. El colector de T1 tiene un valor de tensión próximo a cero. En este estado, T2 recibe una polarización baja y no conduce ,dando salida alta (salida 2), mientras que la salida 1, colector de T1 es baja. Este estado puede considerarse **de reposo**.

En la misma figura están representados los cuatro valores de tensión en el colector de T1, base y colector de T2 y base de T1, cuando llega un impulso negativo a la base de T1, que lo corta y cambia el estado. Veamos con todo detalle como ocurre el proceso siguiente:

El colector de T1 sube desde su valor de saturación (despreciable, aunque indicado en la figura) a un valor próximo a E_b.

La base de T2 recibe una polarización a través de la resistencia de base R_{b2} y el transistor T2 conduce.

El colector de T2 baja bruscamente (al conducir T2) desde el valor E_b hasta el pequeño valor de saturación V_{SAT}, que es muy bajo.

El **condensador C** transmite íntegramente la bajada de tensión de valor $E_b - V_{SAT}$, de forma que la tensión en la base de T1 baja al valor $V_{B1} - E_b + V_{SAT}$.

La tensión en la base de T1 pasará exponencialmente desde el valor inicial hasta cargarse a $+E_b$ con la constante de tiempo CR_{b1}. La ecuación que da la tensión V(t) en la base de T1 será:

$$V(t)=E_b-(2E_b-V_{B1}-V_{SAT})\,e^{-\frac{t}{CR_{b1}}}$$

y llegará otra vez al valor V_{B1} en el tiempo t que satisface a la ecuación:

$$E_b-(2E_b-V_{B1}-V_{SAT})\,e^{-\frac{t}{CR_{b1}}}=V_{B1}$$

Si despreciamos los valores de saturación V_{B1} y V_{SAT}, la ecuación queda:

$$E_b-2E_b e^{-\frac{t}{CR_{b1}}}=0$$

$$es\ decir\ e^{-\frac{t}{CR_{b1}}}=0,5$$

$$t=CR_{b1}\ln 2=0,69\,CR_{b1}$$

Este es el valor del tiempo de "recuperación". Cuando ha transcurrido, T1 conduce de nuevo y su colector baja, T2 se corta y su colector sube.

En la salida 1 se ha obtenido un impulso rectangular positivo de la duración t calculada y en la 2 se obtiene un impulso negativo de la misma duración. El circuito es un **temporizador.**

11.9.- Multivibrador libre o astable.

La figura 11-19 representa un multivibrador libre, que oscila sin entrada exterior, produciendo una salida **rectangular**, que puede extraerse de uno u otro colector, de la forma que se ve en la misma figura.

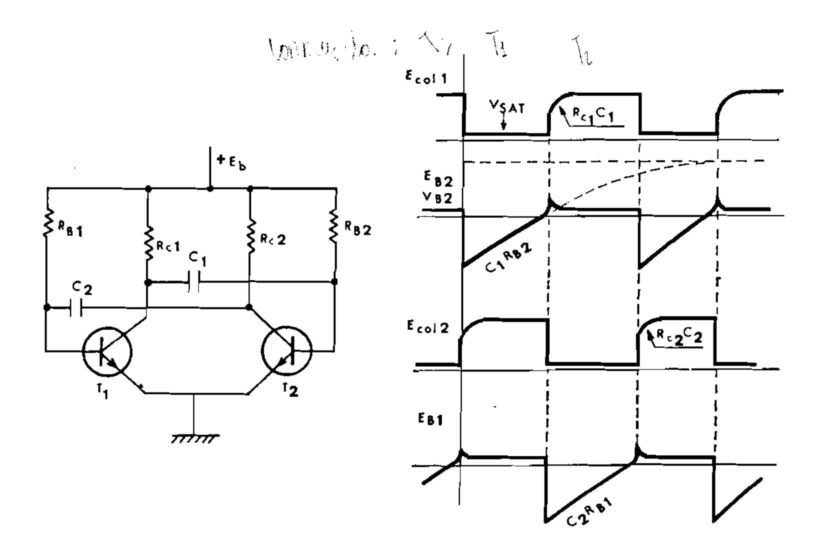


Fig. 11-19.

El circuito está basado en el mismo principio de realimentación a través de un condensador del multivibrador monoestable, pero en ambos sentidos, con dos condensadores. La fuerte realimentación positiva hace que sean inestables todos los estados intermedios y el acoplamiento CR fija la duración de cada estado, lo mismo que ocurría con el circuito anterior, pero para ambos estados. A la derecha de la figura se ven las tensiones en colectores y bases en función del tiempo.

Podemos comenzar el proceso suponiendo que, por motivos que se verán después, T1 comienza a conducir. Como se ve en la figura, eso significa que su colector baja bruscamente de E_b a V_{SAT}, es decir, baja un valor muy próximo a E_b.

La base de T2 recibe este salto negativo a través del condensador C_1, con lo que T2 se corta. El colector de T2 sube a la tensión E_b, cargando el condensador C_2 a través de R_{C2} y la base de T1, lo que se nota por la subida exponencial de la figura. La constante de tiempo de los colectores es mucho más baja que la de las bases, por lo que este retardo no influye en el funcionamiento del multivibrador. La base de T1 conserva una tensión de polarización suficiente para la saturación, que recibe a través de R_{B1}.

Este estado (T1 conduciendo y T2 no conduciendo) se conserva hasta que la base de T2 (segunda curva) se recupere exponencialmente. Como hemos visto en el párrafo anterior, esto ocurre cuando ha transcurrido el tiempo 0,69 C_1 R_{B2}. Entonces T2 empieza a conducir y transmite a través de C_2 un "frente" o cambio negativo de valor aproximado E_b a la base de T1.

Ahora es la base de T1 la que se "recupera" exponencialmente

en el tiempo 0,69 C_2 R_{B1}. Cuando transcurre ese tiempo, T1 vuelve a conducir y todo el proceso se inicia de nuevo como hemos indicado al principio.

Se trata pues de un circuito en "balancín" o "báscula" . Las transiciones son muy rápidas, por lo que se le llama también **oscilador de relajación.** Los tiempos de duración de cada estado son 0,69 C_1 R_{B2} y 0,69 C_2 R_{B1}, que pueden ser distintos, dando ondas rectangulares en las dos salidas que son los colectores, siendo una onda la inversa de la otra.

Lo más corriente es que se haga el circuito totalmente simétrico, o sea R_{C1} = R_{C2}, R_{B1} = R_{B2} y C_1 = C_2, con ambos transistores idénticos. Se dice entonces que produce una **onda cuadrada** (la forma sigue siendo rectangular pero ambos estados de la misma duración). El circuito es un **reloj**.

11.10.- Circuito Schmitt.

El llamado **circuito Schmitt o Schmitt-trigger** es un caso particular de multivibrador. Es un circuito biestable en el que el paso de uno a otro estado ocurre en dos **umbrales** de la tensión de entrada. Ambos umbrales están cruzados con una curva de **histéresis** que es muy característica de este circuito y que veremos a continuación.

El circuito de la figura 11-20 es un Schmitt-trigger. Despreciando las tensiones de saturación, hay dos posiciones estables, claramente asimétricas (obsérvese que la resistencia de emisor forma parte de la realimentación):

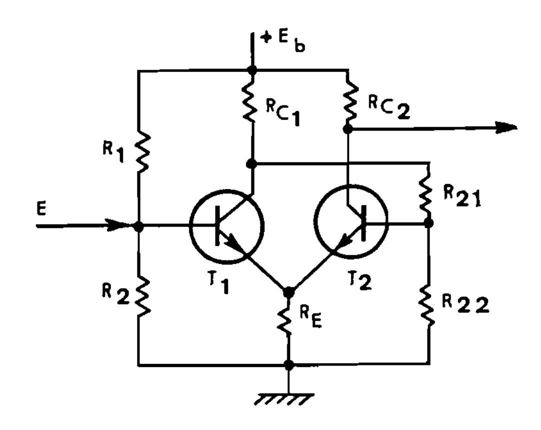


Fig. 11-20.

a) Si conduce T1, la tensión en base y emisor de T1 es:

$$V_1 = E_b \frac{R_E}{R_E + R_{C1}}$$

Se supone R_{C1} mucho menor que R_{21} y R_{22}. La tensión de emisor T2 la suponemos igual a la de base (despreciamos también las tensiones de base en conducción), V_1. La tensión del colector de T1 también tiene el mismo valor V_1. La tensión en la base de T2 es inferior, por lo que T2 **no** conduce.

Para mantener este estado en conducción se elige:

$$\frac{R_2}{R_1+R_2} > \frac{R_E}{R_E+R_{C1}}$$

para que T1 esté polarizado en conducción y el estado permanezca así hasta que se cambie desde fuera.

b) Si bajamos la tensión de entrada E hasta un valor inferior a V_1, T1 se corta. El divisor R_{21}, R_{22} se calcula para que T2 conduzca (basta elegir R_{22} alto). En ese momento, habiendo elegido $R_{C2} < R_{C1}$, la tensión en el emisor **sube**, ayudando al corte de T1, al valor:

$$V_2 = E_b \frac{R_E}{R_E+R_{C2}}$$

R_1 y R_2 se calculan para que, ahora, con el nuevo valor de la tensión de emisor, T1 permanezca cortado, es decir, la doble condición del divisor de tensión de entrada es:

$$\frac{R_E}{R_E+R_{C1}} < \frac{R_2}{R_1+R_2} < \frac{R_E}{R_E+R_{C2}}$$

Si ahora se quiere volver al estado a hay que elevar E hasta pasar el segundo umbral V_2.

La figura 11-21 representa la salida E_{C2} en función de la entrada E, resaltando el carácter de **histéresis** de la dependencia.

El Scmitt-trigger se aplica para convertir formas irregulares en ondas rectangulares bien definidas, o para conmutar sin permitir valores intermedios. La figura 11-22 es un ejemplo de conversión de una función de forma irregular en una onda del tipo rectangular, con dos valores discretos a los que se llega cuando la entrada sobrepasa en uno y otro sentido los umbrales de conmutación.

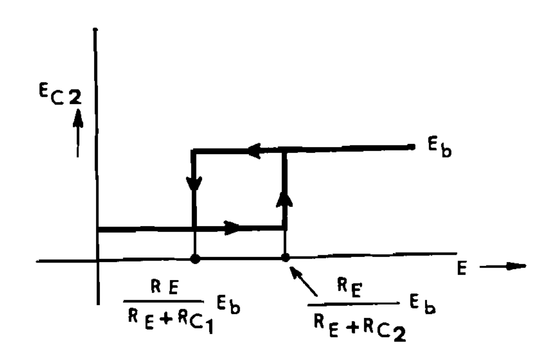


Fig. 11-21.

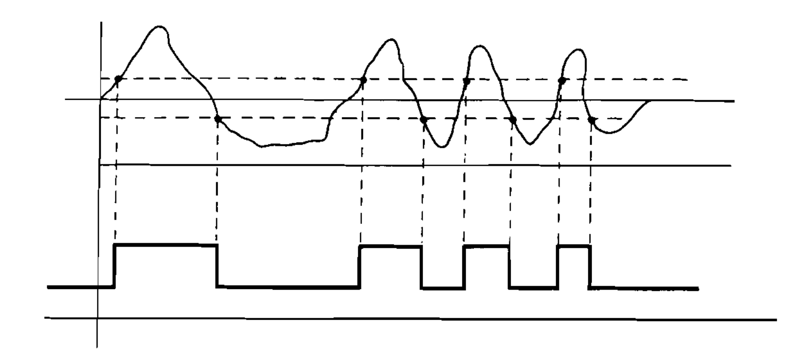


Fig. 11-22.

11.11.- Flip-flop con transistores MOSFET.

El flip-flop estudiado en el párrafo 11-7 es un conjunto de dos inversores conectados entre sí de forma que la salida de cada uno de ellos controla la entrada del otro. El mismo tipo de circuito puede construirse con toda facilidad con otros tipos de conmutadores. Así por ejemplo, el circuito de la figura 11-23 es un flip-flop con transistores complementarios del tipo CMOS.

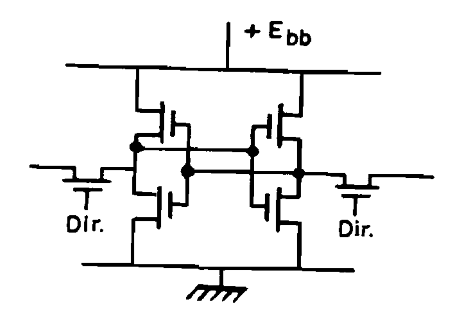


Fig. 11-23.

Además de su enorme sencillez y de no necesitar resistencias, este circuito puede completarse con dos circuitos de lectura-escritura que permiten aislarlo del exterior o conectarlo al exterior cuando se activen estos conmutadores. La señal que abre estos conmutadores se llama **direccionamiento** y permite dirigirse a un flip-flop particular dentro de un conjunto de circuitos del mismo tipo que constituyen una memoria.

11.12.- El amplificador operacional como conmutador.

Lo mismo que, como vimos, el transistor se convertía en un conmutador usándolo en posiciones extremas, el amplificador operacional puede usarse en sus dos posiciones extremas o valores de saturación, aplicando señales de entrada mayores de las normales.

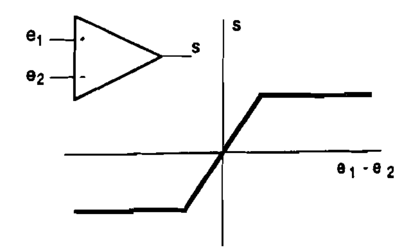


Fig. 11-24.

La figura 11-24 representa la tensión de salida en función de la de entrada en un amplificador operacional diferencial, incluyendo las zonas de saturación. La amplificación A se supone muy grande, lo que hace que la zona de trabajo s/A sea muy estrecha. Despreciando esta zona, el amplificador es un conmutador tal que si $e_1 - e_2$ es negativo, la salida tiene un valor negativo alto, próximo a la alimentación negativa V-. Si la entrada es positiva, la salida tiene un valor alto positivo, próximo a la alimentación positiva V+. Además de conmutar entre estos dos valores, el amplificador operacional es un **comparador,** que da una salida u otra según que $e_1 > e_2$ o que $e_1 < e_2$.

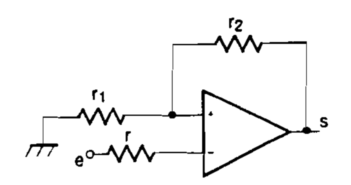


Fig. 11-25.

A diferencia con el transistor, la amplificación es positiva con la entrada adecuada, y puede obtenerse una realimentación positiva sin necesidad de duplicar el dispositivo. La figura 11-25 es un ejemplo de Schmitt-trigger con amplificador operacional. La realimentación positiva mantiene al amplificador en un lado y el valor de entrada positiva es una fracción de la salida. Hasta que la entrada negativa sobrepase el **umbral** así obtenido no conmuta el emplificador, que coloca la entrada positiva en el umbral opuesto.

Conectando un RC a la entrada negativa de la figura anterior, se obtiene un multivibrador libre, ya que cada vez que conmuta el amplificador comienza a cargar el condensador hasta que llega al umbral del mismo signo, momento en que invierte el proceso. La figura 11-26 es un multivibrador de este tipo (un reloj). En la misma figura pueden verse las curvas que definen el comportamiento del circuito.

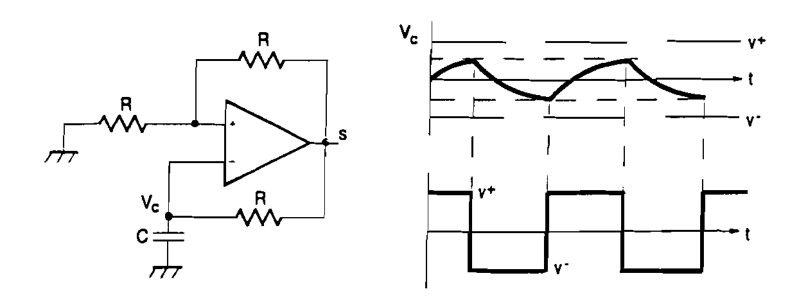


Fig. 11-26.

11. 13.- Circuitos lógicos "o" e "y" con diodos.

Vamos a estudiar los circuitos clásicos de conmutación con diodos con señales cualesquiera de entrada, para particularizarlos luego al caso de aplicaciones digitales.

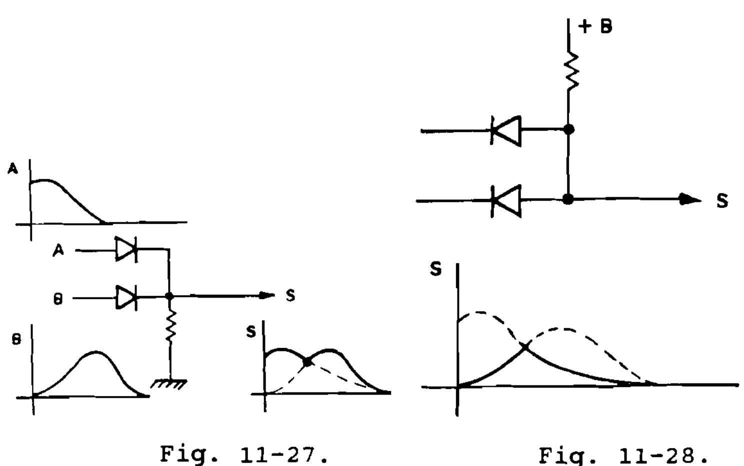


Fig. 11-27. Fig. 11-28.

El circuito de la figura 11-27 es un circuito **o** con diodos. Si las entradas A y B son analógicas, la salida es la **mayor** de las entradas y sólo es cero cuando ambas son cero o inferiores. En efecto, si A > B, conduce el diodo de entrada A y la salida es igual a A, S=A, con lo que el diodo de la entrada B no conduce. Del mismo modo, si A < B, conduce el diodo de la entrada B y la salida es S = B, con lo que no conduce el otro diodo. Si se conectasen más diodos a la resistencia, el resultado sería análogo, conduciendo el que tenga mayor entrada. La curva S dibujada explica gráficamente el mismo resultado, S es igual a la mayor de las entradas.

La figura 11-28 es un circuito **y**. La salida es la **menor** de las entradas, que es la que hace conducir al diodo correspondiente. Los demás diodos quedan con polarización inversa, al aparecer a la salida un valor menor que todos los demás (que pueden ser más de dos).

La salida del circuito **o** de la figura 11-27 corresponde a la operación lógica **o**, **or**. Si las entradas sólo toman dos valores que denominamos 0 y 1, se puede definir una **tabla de correspondencia** de la función S de A y B que también se llama **suma** (lógica) o disjunción. La tabla es:

A	B	S
0	0	0
0	1	1
1	0	1
1	1	1

También puede escribirse la tabla con el convenio **L** (bajo, 0) o **H** (alto), en lugar de **0** y **1**, poniéndose de manifiesto que la denominación lógica puede ser variada. La tabla anterior define la **función suma lógica** o **función de Boole S = A + B.**

Del mismo modo, el circuito de la figura 11-28 corresponde a la operación lógica **y**, **and**. Si las entradas toman los valores 0 y 1, que podemos identificar con las tensiones 0 y alimentación, se puede definir una **tabla de correspondencia** de la salida S en función de A y B. Esta función es el **producto lógico** o **función de Boole S = A * B**, o, simplemente, **S = AB.**

A	B	S
0	0	0
0	1	0
1	0	0
1	1	1

Hay que hacer notar que el producto lógico también expresa

la **coincidencia** de las entradas, ya que la salida es 1 sólo cuando ambas entradas son 1.

Los circuitos digitales anteriores se llaman también en terminología digital **puertas lógicas**. En diagramas de bloques degitales, se representan las puertas como se indica en la figura 11-29. Los símbolos de la izquierda son los más usados, corresponden a la norma americana, mientras que los de la derecha corresponden a la norma DIN alemana.

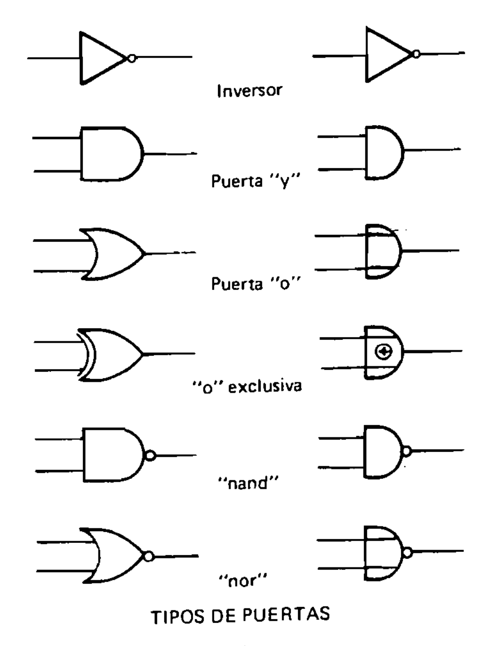


Fig. 11-29.

En esta figura se han incluido también circuitos **inversores** (los primeros), ya estudiados en conmutación, que materializan la **negación lógica**, $S = \bar{E}$, que se lee **S = no E**.

Los pequeños círculos representados a la salida de algunas de estas **puertas** representan que se ha conectado un inversor que hace la negación de lo que salga. Así, con esa negación, la puerta y se convierte en **NAND**, (no y) y la puerta o se convierte en **NOR** (ni). Un caso especial lo constituye la puerta **o exclusiva**, cuya tabla de correspondencia es la que sigue:

A	B	S
0	0	0
0	1	1
1	0	1
1	1	0

Se ve que corresponde a un "o" distinto del anterior, ya que expresa que el valor de A o B es 1 pero no ambos. Corresponde al "aut" latino (en lugar del "vel" de la lógica o). Para realizar esta puerta hay que combinar las demás en la forma que estudiaremos en un capítulo posterior.

11.14.- Familias de circuitos integrados.

La aparición de técnicas de integración cada vez más completas para integrar en una sola pastilla o "chip" gran cantidad de componentes ha dado lugar a la aparición de tecnologías normalizadas para la fabricación de las puertas digitales antes descritas, llamadas también **familias de circuitos integrados**. Una familia es un conjunto de circuitos integrados que utilizan una tecnología común.

Las más importantes son:

a) Familia **RTL** (Lógica de resistencias y transistores).

Se basan en la utilización de inversores de transistores bipolares, con varias resistencias de entrada para obtener el efecto lógico deseado. Por ejemplo, la figura 11-30 es un circuito **NOR**. Si cualquiera de las entradas es alta, el transistor conduce y la salida es baja. La ecuación de Boole es:

$$S=\overline{A+B}$$

Al conjunto de entradas se le llama **abanico de entrada** (fan in) y al número de circuitos que puede conectarse a la salida (en este caso depende de la resistencia de colector, pero serán unos 10), se le llama **abanico de salida** (fan out).

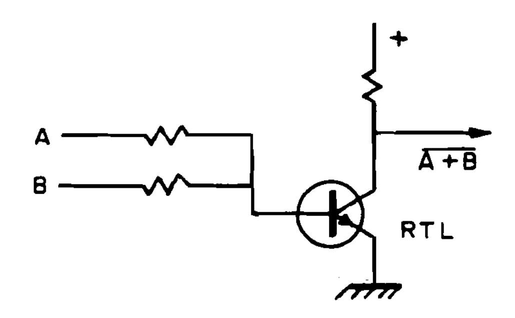


Fig. 11-30.

La figura 11-31 es otro circuito de tipo **NOR** con dos transistores. Cuando alguna de las entradas es alta, el transistor correspondiente conduce y la salida es baja.

Fig. 11-31.

b) Familia **DTL** (lógica de diodos y transistores).

Se trata de los mismos circuitos de diodos "y" y "o" que hemos estudiado ya, seguidos de un inversor. Así, la figura 11-32 es un circuito **NAND** compuesto de esa forma.

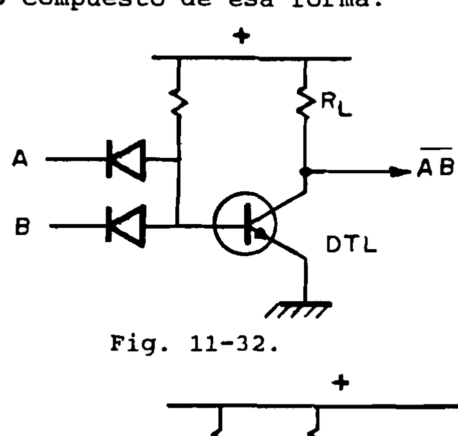


Fig. 11-32.

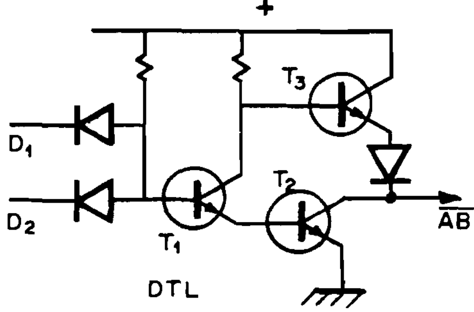


Fig. 11-33.

Normalmente se añade un seguidor de emisor, de baja impedancia en alto, complementado por otra etapa de salida que asegure baja impedancia "en bajo". El circuito de la figura 11-33 representa un circuito NAND con salida llamada **en cascada**, o también en **totem.** El circuito lógico es el mismo de la figura anterior, con dos diodos que hacen la función lógica "y", seguidos de una etapa inversora, T1. Cuando T1 conduce, lo hace también T2 y la salida es baja y de baja impedancia. Cuando T1 no conduce, por ser baja alguna de las entradas, su colector está

alto y T3 hace ahora de seguidor de emisor dando salida alta de impedancia baja. El diodo garantiza una polarización mínima al emisor de T3 y, por consiguiente, al colector de T1, aunque se cortocircuite la salida.

En algunos casos se utiliza el circuito de la figura 11-32 sin la carga de colector R_L, dejando la carga en el exterior. Entonces se le llama **en colector abierto**.

c) Circuitos **TTL** (lógica de transistor a transistor).

Debido a la facilidad de integración de los transistores, a veces se prefiere sustituir los diodos por transistores multielectrodos. Como se ve en la figura 11-34 se sustituyen los diodos de entrada por un transistor multiemisor. Por lo demás, el circuito es exactamente igual al de la figura 11-33.

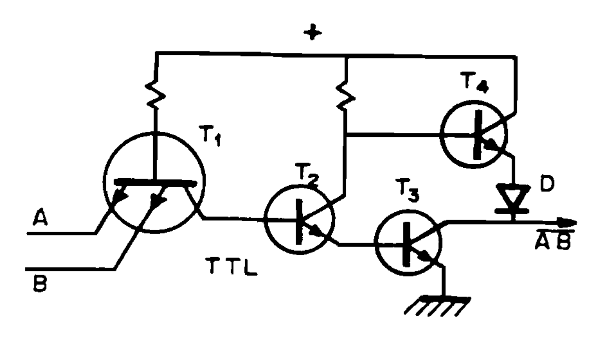


Fig. 11-34.

d) Circuitos **ECL**. (lógica de acoplamiento de emisor).

Se utiliza un circuito diferencia de entrada con lógica de transistores, como se ve en la figura 11-35.

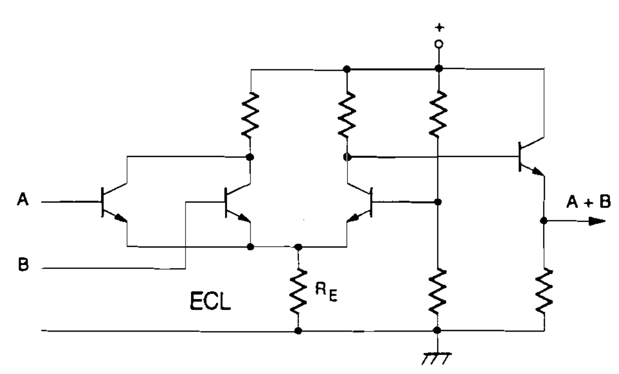


Fig. 11-35.

Cuando una de las entradas A o B es alta, cortocircuita la parte izquierda del amplificador diferencial y corta la derecha (por dar más tensión al emisor común). La salida, a través de un seguidor de emisor, es alta, es decir, el circuito es **or**. Las

polarizaciones están estudiadas sin saturación, lo que hace la conmutación más rápida. Es la familia más rápida de conmutación con transistores.

Existen otras familias de circuitos integrados, caracterizadas por los respectivos inversores, como MOSFET, CMOS, Schottky, etc... y otras que son variantes de las anteriores como la I^2L (lógica de doble inyección), que acelera la conmutación con un transistor complementario, etc...

CAPITULO 12

FUENTES DE ALIMENTACION

12.1.- Generalidades.

Todos los sistemas electrónicos necesitan una o varias fuente de energía para poder funcionar. Salvo alguna excepción, la alimentación necesaria es de corriente continua. Hasta ahora hemos representado las fuentes siempre en forma de "baterías" o generadores ideales de corriente continua, pero normalmente se obtiene la energía a partir de una fuente única que puede ser una batería o la red de alimentación de energía eléctrica.

En el caso de alimentación real por batería o baterías será necesario disponer de las pilas o acumuladores que se necesiten, excepto las alimentaciones que quedan hacerse a través de divisores de tensión. Debido al mal rendimiento de estos últimos, su uso estará limitado todo lo posible y siempre en partes del sistema que necesiten muy poca potencia. La transformación de un valor a otro de c.c. es complicada y se suele evitar en lo posible.

La alimentación más cómoda es la que se obtiene a partir de la red de c.a. Con transformadores pueden obtenerse todos los valores que se necesiten y con rectificadores se pueden transformar en c.c. con facilidad.

También hay que tener en cuenta que en los sistemas electrónicos se necesitan muchas veces fuentes de alimentación de gran precisión, muy superior a la precisión normal del

suministro de energía eléctrica. Para ello se utilizan **fuentes de alimentación estabilizadas**, que aumentan la precisión de la alimentación a lo que se necesite en cada caso. En este capítulo repasaremos brevemente los principios de transformación y rectificación para estudiar después la estabilización electrónica.

12.2.- Transformadores.

Los transformadores que se utilizan en las fuentes de alimentación son de acoplamiento fuerte (m=1) y de pocas pérdidas, lo que se evita con núcleo de chapa de hierro "magnético", de baja histéresis, con un ligero contenido de níquel, aislada (para disminuir las corrientes de Foucault). Generalmente se usa un transformador único para varias fuentes, con un primario, alimentado con la red, y varios secundarios que suministren los valores deseados.

Normalmente será necesario proyectar los transformadores para la frecuencia de red disponible, 50 Hz. en Europa y 60Hz. en E.E.U.U., pero como la sección necesaria decrece con la frecuencia, a veces, se elige la frecuencia (400Hz. en los aviones) para que los transformadores resulten más ligeros.

Para el proyecto de transformadores remitimos al lector a la literatura de Electrotecnia adecuada (por ejemplo, al libro del profesor Sanjurjo Navarro).

12.3.- Rectificadores.

Para transformar la corriente alterna en continua se utiliza la propiedad de los diodos de conducir en una sola dirección. Durante el tiempo de conducción se carga un condensador, que acumula una carga eléctrica que a su vez suministra a la carga durante todo el tiempo, como un depósito regulador, manteniendo la tensión casi constante. La figura 12-1 representa un rectificador con un condensador C y una carga R y en la figura 12-2 se ha sustituido el diodo por un diodo ideal y una resistencia interna r. Sea E_E la tensión de entrada (alterna) y E_S la de salida obtenida.

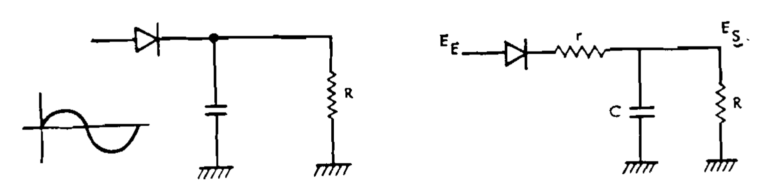


Fig. 12-1. Fig. 12-2.

La tensión de entrada es $E_E = E_m$ sen ωt. La salida (la del condensador) es aproximadamente continua, aunque crece ligeramente durante el período de carga y decrece durante el resto del tiempo, como se aprecia en la figura 12-3.

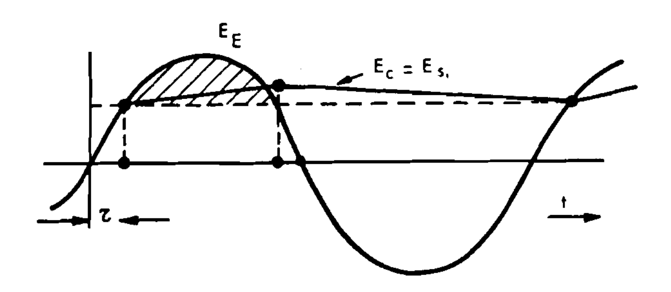


Fig. 12-3.

Vamos a calcular la tensión de salida $E_C = E_S$ en función de la tensión de entrada $E_E = E_m$ sen ωt, con $\omega = 2\ \pi f$ y $f= 1/T$, y de los parámetros del circuito R y r. Este tipo de rectificador, en que sólo se usa un semiciclo de la c.a. se llama de **media onda**.

A la vista de la figura 12-3, puede verse que el tiempo de carga del condensador empieza cuando la tensión de entrada es superior a la del condensador. Despreciando la caída de tensión en el diodo y las variaciones de la tensión de carga, ese momento τ ocurre cuando es $E_E = E_m$ sen $\tau = E_C$. Durante todo el tiempo entre τ y $T/2 - \tau$ está cargando el condensador. Llamemos Q a la carga en culombios que entra en el condensador. Por otra parte, el condensador está descargando todo el tiempo hacia la carga, con una intensidad E_C/R, con lo que la carga que sale del condensador durante todo el período T será TE_C/R. El rectificador está en equilibrio cuando la carga que entra sea igual a la que sale, es decir:

$$Q=\int_{\tau}^{\frac{T}{2}-\tau}\frac{E_m sen\omega t-E_m sen\omega\tau}{r}dt=\frac{E_m sen\omega\tau}{R}T \qquad (1)$$

es decir:

$$\frac{1}{\omega r}\int_{\omega\tau}^{\pi-\omega\tau} sen\omega t d(\omega t)-[\frac{1}{r}sen\omega\tau]\,[\frac{T}{2}-2\tau]=\frac{T}{R}sen\omega\tau$$

$$\frac{2}{\omega r}\cos\omega\tau=sen\omega\tau\,[T(\frac{1}{2r}+\frac{1}{R})-\frac{2\tau}{r}]$$

Llamando $\omega\ \tau = \varphi$ y teniendo en cuenta que $\omega T = 2\ \pi$,

$$ctg\varphi = \frac{\pi}{2} + \frac{\pi r}{R} - \varphi$$

ecuación transcendente en φ que, resuelta por algún procedimiento gráfico o aproximado da el valor de la fase y, por consiguiente, de E_C en función de E_m, ω y r/R. Normalmente se tabula el valor r/R en función de φ (radianes) y se obtiene así una tabla que permite leer los valores que se deseen. Por ejemplo, para r/R = 0,05 resulta ser la tensión continua de salida aproximadamente igual a la eficaz de entrada, condición que se utiliza frecuentemente en el proyecto de rectificadores.

El rectificador descrito y calculado es el de media onda, porque carga el condensador sólo durante una fracción del semiciclo positivo. El circuito de la figura 12-4 es un **rectificador de onda completa.** Mediante un devanado inversor de fase en el transformador se alimenta un segundo diodo de forma que el condensador se carga el **doble** del tiempo que en el anterior.

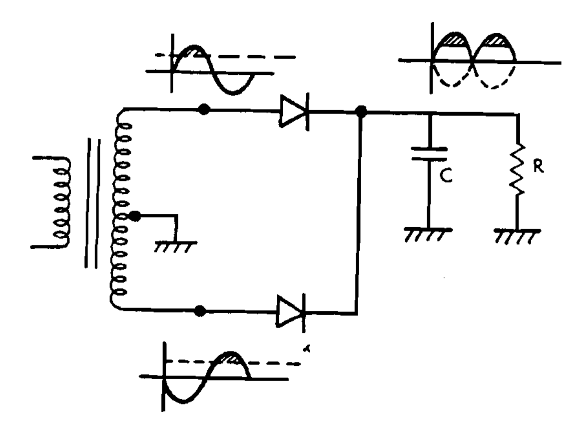


Fig. 12-4.

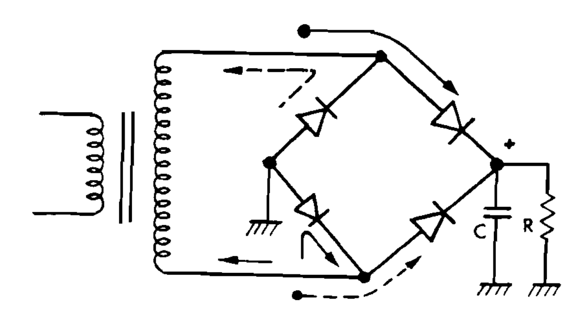


Fig. 12-5.

La figura 12-5 obtiene el mismo resultado mediante un circuito con cuatro diodos y un solo secundario del transformador. Durante medio ciclo, el condensador se carga por el trayecto indicado con línea completa y durante el otro medio por el trayecto indicado a trazos. La pequeña variación de la tensión del condensador, llamada **ondulación residual** tendrá doble frecuencia que la de la red. Este circuito se llama **puente de Graetz**.

La ecuación que define la fase de carga es ahora:

$$Q=2\int_{\tau}^{\frac{T}{2}-\tau}\frac{E_m sen\omega t-E_m sen\omega\tau}{r}dt=\frac{TE_m sen\omega\tau}{R}$$

El efecto es el mismo que si en la ecuación anterior se hubiese hecho la carga con resistencia interna r/2, lo cual es favorable, o, también, si la carga hubiese sido 2R. Es decir, que puede decirse que el rectificador maneja el doble de carga (mitad de R) en las mismas condiciones que el de media onda. Por ejemplo, ahora la tensión de salida será aproximadamente igual a la eficaz para r = 0,1 R.

Por último, veamos cuanto vale la **ondulación residual**, en ambas clases de filtros. Para hallarla, se calcula cuanto varía carga del condensador, es decir, la cantidad de electricidad que entra en el condensador durante el período de carga (aumentando E_C) y la que sale durante el resto. Esta carga no es Q de las fórmulas anteriores, que era la carga que entraba **hacia** el condensador, pero una parte iba a la carga exterior. Es más fácil calcular la carga que sale del condensador cuando no está cargando.

En el rectificador de media onda, la carga que sale del condensador entre el tiempo $T/2 - \tau$ y $T + \tau$ vale:

$$Q_C=\frac{E_C}{R}\left(\frac{T}{2}+2\tau\right)$$

Suponemos que durante todo el tiempo E_C no varía apreciablemente. La variación (pequeña) de E_C vale

$$\Delta E_C=\frac{Q_C}{C}=\frac{E_C}{CR}\left[\frac{T}{2}+2\tau\right]$$

La ondulación es creciente con τ. Esta ondulación es una onda aproximadamente triangular de frecuencia la de la red. La

ondulación es el valor pico a pico (entre máximo y mínimo). El caso más desfavorable sería (ver figura 12-3) con τ = T/4, y la ondulación en ese caso sería E_C T / CR. La ondulación **relativa** vale:

$$\frac{\Delta E_C}{E_C}=\frac{1}{CR}\left[\frac{T}{2}+2\tau\right]<\frac{T}{CR}$$

CR es la constante de tiempo del condensador C con la carga R. Siempre en el caso más desfavorable, la fracción T/CR es la ondulación relativa. Si, por ejemplo, se desea una ondulación inferior al 5%, (0,05), la constante CR debe ser unas 20 veces el período. La ondulación obtenida será algo menor, debido a que τ no llega a valer T/4.

Si se supone ahora un valor conservador de proyecto de τ = T/8, la ondulación pasa a ser:

$$\frac{\Delta E_C}{E_C}=\frac{3}{4}\frac{T}{CR}$$

y, para el mismo ejemplo anterior de ondulación relativa 5%, la constante CR debe ser 15 veces el período. (Para 50 Hz., el período es 20 mS y CR debe ser 0,3 segundos).

En el caso del rectificador de onda completa, la ondulación relativa vale:

$$\frac{\Delta E_C}{E_C}=\frac{E_C}{CR}4\tau$$

El caso más desfavorable de τ = T/4 da el mismo resultado del caso anterior, pero el caso de proyecto más normal con un valor τ= T/8 da una ondulación relativa :

$$\frac{\Delta E_C}{E_C}=\frac{1}{2}\frac{T}{CR}$$

y, para el mismo ejemplo de ondulación relativa 5%, CR debe ser 10 veces el período. Parece que la ventaja no ha sido tan grande al usar rectificación completa (CR de 15 a 10 veces T), pero no hay que olvidar que, además, para el mismo valor de τ, se podía cargar el doble, dividir R por 2.

12.4.- Filtros.

La ondulación residual es un problema importante en los circuitos electrónicos y es necesario reducirla todo lo posible. Para reducirla se intercalan filtros pasabajos entre el rectificador propiamente dicho y la carga. La figura 12-6 representa un filtro LC del tipo indicado.

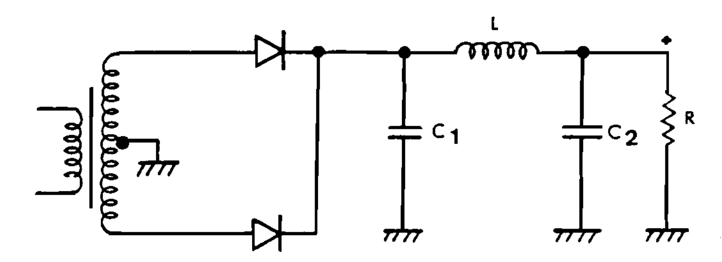


Fig. 12-6.

El rectificador con el primer condensador puede tratarse como un rectificador de onda completa análogo al de la figura 12-4 con C_1 como condensador y R como carga, ya que la corriente a través de la autoinducción tiende a ser aproximadamente constante. Así se obtendrá el valor de E_C y de la ondulación ΔE_C. Ahora puede tratarse el filtro como un problema lineal con una entrada con una componente de continua E_C, que pasa inalterada por el LC, y una componente alterna de valor pico a pico ΔE_C y frecuencia 2f, doble de la de la red.

El filtro se calcula con una frecuencia propia de resonancia muy inferior a la de la red, es decir:

$$\frac{1}{2\pi\sqrt{LC}} < 2f$$

o, lo que es lo mismo, se hace que la impedancia de C_2 sea muy baja y la de L muy alta a la frecuencia de la ondulación, con lo que la amplificación (atenuación) a esa frecuencia vale:

$$\frac{\frac{1}{\omega C}}{\frac{1}{\omega C}+\omega L} = \frac{1}{1+\omega^2 CL} = \frac{1}{1+16\pi^2 f^2 CL} \approx \frac{1}{16\pi^2 f^2 CL}$$

La ondulación triangular tiene componentes de frecuencias superiores, que son mucho más atenuadas, de forma inversamente proporcional al cuadrado de la frecuencia. Por consiguiente puede suponerse que solo permanece una ondulación de valor pico a pico que se obtiene multiplicando ΔE_C por la atenuación obtenida.

12.5.- Estabilización con diodo Zener.

Las fuentes de alimentación descritas en los párrafos anteriores tienen los siguientes inconvenientes:

La red de suministro de energía eléctrica tiene variaciones de ± 15%, que se transmiten a la salida. Además, los errores o pequeños defectos de los componentes producen una desviación adicional en el valor de salida.

Por otra parte, la impedancia de salida no es despreciable (ya hemos visto la dependencia de la tensión de salida en función de la resistencia de carga) o, lo que es lo mismo, la tensión de salida depende demasiado de la intensidad de salida.

Las fuentes que de alguna forma compensan estas variaciones de la tensión de entrada y de la carga se llaman **estabilizadas**. Para estabilizar la tensión de salida hay que tomar una **referencia de tensión** constante, que, en la gran mayoría de los casos es un **diodo Zener**.

Una primera clase de fuentes estabilizadas con Zener es la que utiliza directamente el valor Zener como valor estabilizado. Estas fuentes se han hecho muy populares debido a su sencillez y bajo precio y son las que se suelen denominar como estabilizadas con Zener, aunque, como veremos después, las demás también utilizan un diodo Zener como tensión de referencia.

La estabilización directa con diodo Zener se consigue con un circuito como el de la figura 12-7.

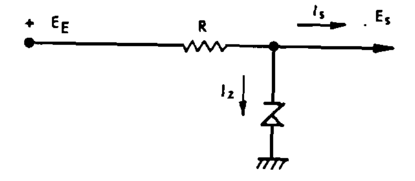


Fig. 12-7.

Las ecuaciones de este circuito son (válidas si $I_Z > 0$):

$$E_S = E_Z + I_Z r_Z \text{(Ecuación del Zener)}$$
$$E_E - E_S = R(I_S + I_Z) \text{ (Ecuación de carga)}$$

Eliminando I_Z se obtiene la ecuación de salida del cuadripolo:

$$E_S = E_E \frac{r_Z}{R + r_Z} + E_Z \frac{R}{R + r_Z} - I_S \frac{R r_Z}{R + r_Z}$$

La tensión de salida es aproximadamente E_Z si suponemos r_Z despreciable frente a R. El término en E_Z es constante, el error está representado por los otros dos términos.

El primer término nos muestra que las variaciones de la tensión de entrada se reflejan ahora a través de un divisor de tensión, que las multiplica por $r_Z/(r_Z + R)$. El último término nos dice que la impedancia de salida del circuito es el resultado de poner en paralelo R y r_Z, es decir, aproximadamente r_Z. Para que las ecuaciones sean válidas, debe ser $I_Z > 0$.

El valor de I_Z se obtiene de la ecuación del diodo y es:

$$I_Z=\frac{E_S-E_Z}{r_Z}=\frac{E_E}{R+r_Z}-\frac{E_Z}{R+r_Z}-I_S\frac{R}{R+r_Z}$$

y se hace cero cuando $I_S = (E_E - E_Z)/R$. El gráfico que representa E_S en función de I_S será una recta de poca pendiente (la que corresponde a la baja impedancia de salida) hasta el punto en que se anula la intensidad por el diodo. A partir de ese momento, el diodo no conduce y la ecuación pasa a ser:

$$E_S=E_E-RI_S$$

con pendiente -R, sin estabilización. La figura 12-8 representa esta dependencia.

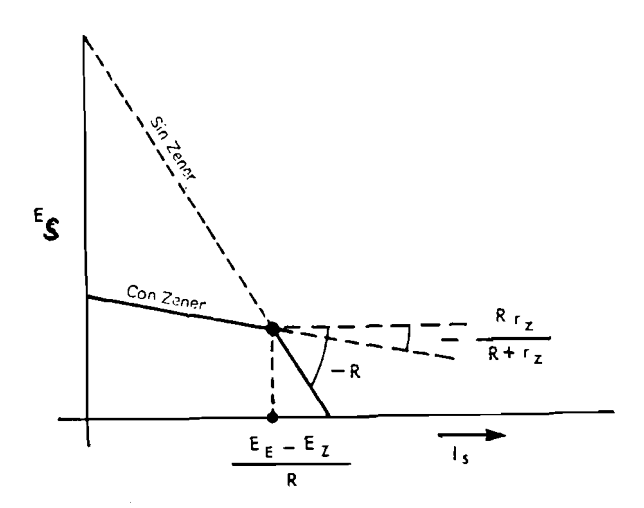


Fig. 12-8.

Las ecuaciones pueden escribirse también en forma

simplificada (teniendo en cuenta que $r_Z<<R$:

$$E_S=E_Z-r_Z I_S$$
$$I_Z=\frac{E_E-E_Z}{R}-I_S$$

R se proyecta con dos condiciones:

Que la intensidad I_Z no baje de un cierto valor I_{Zo} próximo a 0, pero no nulo para evitar la curvatura de la curva de ruptura del Zener, cuando la tensión de entrada sea mínima y la carga máxima.

Que no sobrepase el valor permisible I_Z máx. para que la potencia máxima disipada en el diodo Zener, $W_Z = E_Z\ I_Z$ no exceda del valor térmicamente permisible para la componente, dado por el fabricante. Estas dos condiciones son :

$$I_{Zmáx}=\frac{E_{Emáx}-E_Z}{R}-I_{Smín.}<\frac{W_{Zmáx.}}{E_Z}$$
$$I_{Zmín.}=\frac{E_{Emín.}-E_Z}{R}-I_{Smáx.}>I_{Zo}$$

Estas condiciones imponen un valor a R que cumpla las condiciones:

$$\frac{E_{Emáx.}-E_Z}{I_{Zmáx.}+I_{Smín.}}<R<\frac{E_{Emín.}-E_Z}{I_{Zmín}+I_{Smáx.}}$$

No olvidar que **no** se trata de manejar una variable R, sino de hacer con este parámetro que se cumplan los valores límites de I_Z. Evitar el error de despejar R y buscar valores mínimos y máximos. Es posible que las inecuaciones de R **no** se puedan cumplir porque el primer miembro es superior al tercero, en cuyo caso la estabilización no es posible con ese Zener. En ese caso hay que usar otro de potencia superior para elevar el valor de $I_{Zmáx.}$ y, con ello, rebajar el valor del primer miembro de las inecuaciones.

Por último, hay que elegir R próximo a su valor máximo (el último término) y comprobar que se cumple R >> r_Z, para que la estabilización sea eficaz.

12.6.- Regulación automática.

La estabilización por diodo Zener reduce la resistencia de salida de la fuente a unos ohmios (valor de r_Z) y el efecto de

las variaciones de la red en la relación R/r_Z, que suele ser de 10 a 20. Sin embargo, las tolerancias de fabricación de los Zener suelen ser del 5% y un ajuste de la tensión de salida no es posible.

Si se necesita una mejor precisión, es necesario recurrir a la regulación automática por realimentación. Más adelante estudiaremos los sistemas de regulación automática en general, especialmente los problemas de estabilidad que presenta, pero ahora adelantaremos la descripción de las fuentes de regulación automática, aunque sea de una forma más elemental.

La figura 12-9 es un "diagrama de bloques" de una fuente de alimentación con regulación automática:

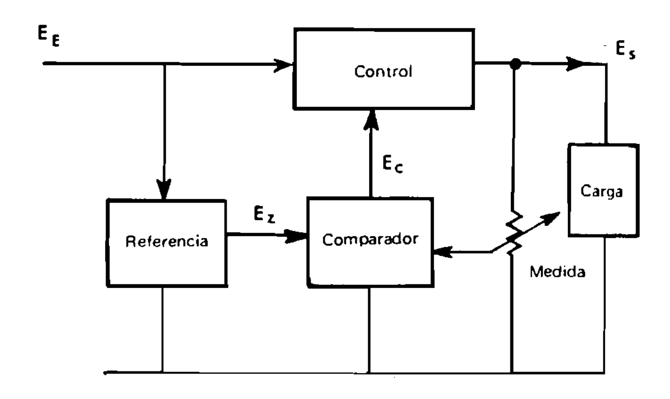


Fig. 12-9.

El núcleo de la regulación es un **comparador** que resta una tensión de referencia de una fracción de la tensión de salida a la que debería ser igual. Si, por ejemplo, la salida debería ser 30 voltios y la referencia (obtenida con un Zener) es 10 voltios, se compara la fracción mE_S de la salida, obtenida con un divisor de tensión, (m= 0,33 en este ejemplo) con la referencia. La diferencia es la **señal de error**, que se amplifica y utiliza para **corregir la salida** en el sentido adecuado, a través del **controlador**. Con este tipo de control (proporcional) se mantiene un ínfimo error para poder amplificarlo y mantener la corrección, pero veremos que ese error es despreciable.

La figura 12-10 es el circuito que materializa la regulación automática en su forma más sencilla. El transistor T1 hace de comparador, ya que mide entre la base (fracción de la salida) y el emisor (valor de referencia del Zener). Su tensión base-emisor es $mE_S - E_Z$, señal de error. Una vez establecido el punto de trabajo, las variaciones de ese valor, es decir, el error respecto a la diferencia $mE_S - (E_Z + V_{BE})$ se amplifica con salida en el colector y se utiliza para controlar al transistor T2 que funciona como un seguidor de emisor.

En una primera aproximación, el sistema está fuertemente

realimentado y, si recordamos los conocimientos de realimentación, **tiende a mantener nula la entrada del amplificador**, es decir, a que $mE_S = E_Z$ (despreciando V_{BE}), con lo que puede **ajustarse** el valor de la salida $E_S = E_Z/m$ variando m, con el potenciómetro de ajuste.

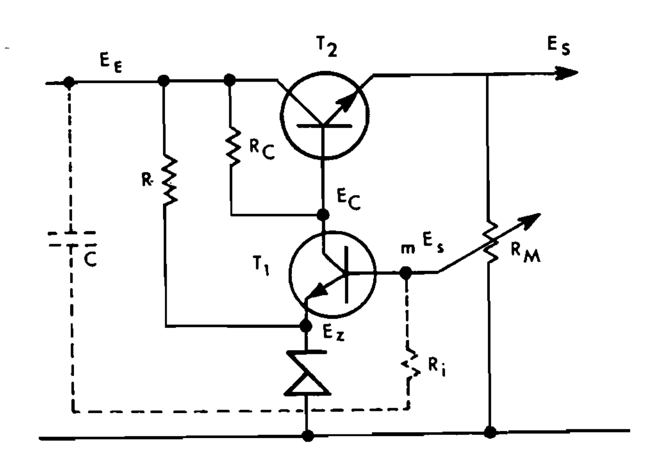


Fig. 12-10.

Una vez establecido el punto de trabajo, podemos calcular sus actuaciones (precisión y resistencia de salida) con las siguientes hipótesis:

La tensión del Zener se mantiene constante (porque su intensidad varía muy poco) y su resistencia diferencial es cero.

El seguidor de emisor T2 se comporta con amplificación unidad y resistencia de salida estimada en r_S.

La amplificación de T1 es $A= - 39\ I_E\ R_C$, despreciando la impedancia de emisor (que sería la del Zener).

Si llamamos con minúsculas las pequeñas desviaciones (debidas a los cambios de la señal de entrada y de la intensidad de salida), las ecuaciones de T1 y T2 son :

$$e_C=-39 I_E R_C m e_S$$
$$e_S=e_C-r_S i_S$$

De ellas se deduce fácilmente:

$$e_s=\frac{1}{1+39\,I_E R_C m}-\frac{r_s}{1+39\,I_E R_C m}i_s\approx\frac{1}{39\,I_E R_C m}-\frac{r_s}{39\,I_E R_C m}i_s$$

donde se ha despreciado el 1 frente a la **amplificación en circuito abierto** $39I_ER_Cm$, que puede tener valores del orden de 200.

De la última fórmula se deduce que las variaciones de entrada y la resistencia de salida **se dividen por la amplificación en circuito abierto**, por ejemplo 200. La resistencia de salida, que ya era baja por ser T2 un seguidor de emisor, queda ahora del orden de centésimas de ohmio en este ejemplo. Compárese con la estabilización por Zener: las variaciones de entrada se dividen por 200 en lugar de 10 y la impedancia de salida queda de, por ejemplo, 0,03 ohmios en lugar de 3 ohmios.

La regulación automática es un sistema de control que puede mejorarse aún más aumentando la amplificación (por ejemplo, con varias etapas en lugar de T1) o "por compensación de la perturbación", si ésta es conocida. Por ejemplo, en la figura 12-10 se ha introducido la ondulación residual de la fuente a través de C y R_i de forma que el transistor T1 le cambia la fase y la añade a la salida en contra de la otra, **compensando** la perturbación.

12.7.- Fuentes de alimentación conmutadas.

Las fuentes de alimentación reguladas estudiadas anteriormente se llaman a veces de "regulación disipativa" ya que hacen la regulación a base de absorber la diferencia de potencia en una resistencia o en un semiconductor. La regulación con diodo Zener disipa una parte importante de la potencia en la resistencia serie y en el Zener y la regulación con amplificador disipa todo el exceso de potencia en el transistor que regula.

A continuación vamos a estudiar otro tipo de regulación que se basa en conmutar intensidades canalizándolas a la carga o a un acumulador de energía, con una distribución de tiempos que es la que hace la regulación. Este es el principio de los reguladores conmutados.

Por último veremos también la generación de corrientes alternas (a partir de continua) mediante conmutación de una intensidad hacia diferentes partes de la carga.

La principal ventaja de las fuentes conmutadas es su mejor rendimiento.

12.8.- Clasificación de las fuentes conmutadas.

Vamos a describir las siguientes.

A. Fuentes de c.c.

A.1. Reguladores con conmutación en serie ("Buck regulators").

A.2. Reguladores con conmutación en paralelo ("Boost regulators").

A.3. Reguladores de transformador de conmutación ("Flyback regulators").

En todos ellos la regulación se hace por ancho de impulso ("Pulse Width Modulation").

B. Fuentes de c.a.

B.1. Onduladores o Convertidores estáticos de frecuencia ("Inverters") no resonantes.

B.2. Onduladores resonantes.

En principio no son reguladores sino simples generadores de frecuencia sincronizados.

12.9.- Reguladores con conmutación en serie (Buck regulator).

El esquema básico con conmutación en serie es el de la figura 12-11.

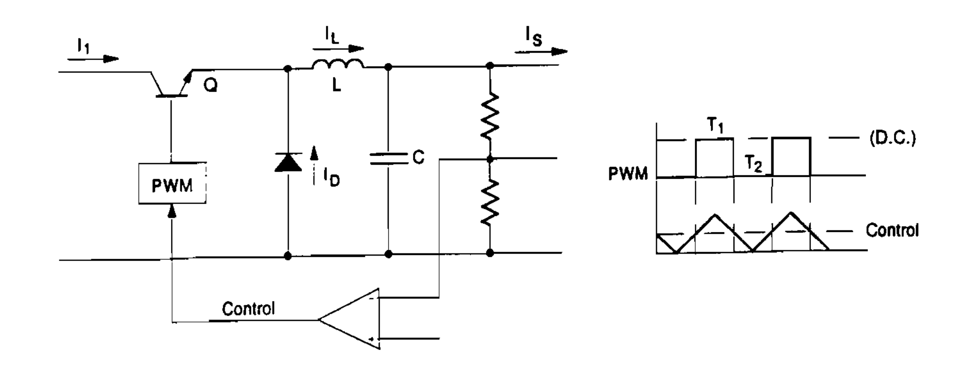


Fig. 12-11.

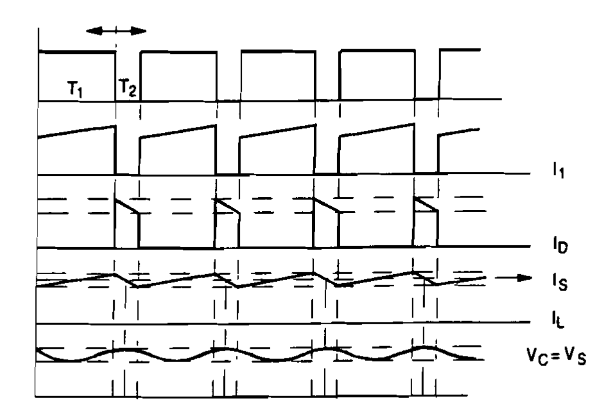


Fig. 12-12.

Veamos como se controla la tensión de salida a través de los impulsos generados en la unidad **PWM** (fig. 12-12). Esta unidad genera una onda rectangular, permaneciendo en "alto" un tiempo T_1 y en "bajo" el tiempo T_2, con suma $T_1 + T_2$ constante, es decir a frecuencia constante. El valor "alto" es el que hace conducir al transistor Q y el valor "bajo" el que lo corta. Como lo importante para la regulación es la proporción del tiempo en que está "alto" el valor en la base de T_1, se podría regular también manteniendo T_1 constante y variando T_2, obteniendo así una "regulación en frecuencia" en lugar de regulación por ancho de impulso.

Suponemos que el condensador C es suficientemente grande para mantener prácticamente constante la tensión de salida V_S (luego veremos que valor debe tener para que las variaciones de V_S sean despreciables). Durante el tiempo T_1 se cortocircuita el transistor Q y la tensión de entrada V_E da lugar a una corriente de carga creciente a través de L. La tensión aplicada a la autoinducción L es $V_E - V_S$ y la ecuación de la autoinducción nos da:

$$V_E - V_S = L\,\frac{dI}{dt}$$

El crecimiento $(\Delta I)_1$ vale:

$$(\Delta I)_1 = \frac{V_E - V_S}{L} T_1$$

Al final del tiempo T_1 se corta el transistor y la corriente I_1 cae a cero. Pero no puede variar bruscamente la intensidad a través de L, que sigue circulando por el diodo D. La inductancia L tiene ahora una tensión cero en el lado del diodo y sigue teniendo V_S a la salida. Su diferencia de potencial es $-V_s$ y la intensidad I_L caerá a un ritmo:

$$-V_S = L \frac{dI}{dt}$$

El decrecimiento durante el tiempo T_2 será:

$$(\Delta I)_2 = -\frac{V_S}{L} T_2$$

y será igual a $-(\Delta I)_1$ cuando se alcance el régimen estacionario.Si el regulador (a través del amplificador) pide más carga, aumenta T_1 y $(\Delta I)_1 - (\Delta I)_2$ es positivo y si el regulador pide menos carga, disminuyendo T_1, esta diferencia se hace negativa. Cuando el amplificador tenga entradas iguales (fracción de la salida igual a la referencia) se llega a un equilibrio y la carga aumenta durante T_1 lo mismo que disminuye en T_2. La intensidad hacia el condensador es I_L y la intensidad que sale hacia la carga (constante si C es suficientemente grande) es I_S, valor central de la rampa de I_L (se desprecia la influencia del divisor de tensión de medida). La diferencia I_L-I_S es la intensidad de carga del condensador y el área dejado por I_L por arriba y por debajo de I_S es la carga (cantidad de electricidad) que entra y sale del condensador en cada ciclo. Esta carga Q vale:

$$Q = \frac{T_1 + T_2}{2} \frac{V_E - V_S}{2L} T_1 = C \Delta V_S$$

de donde se deduce:

$$\Delta V_S = \frac{(T_1 + T_2)\ T_1}{4CL} (V_E - V_S)$$

que es la ondulación de salida. En el caso más desfavorable (T_2 = 0) vale:

$$\Delta V_S = \frac{T_1^2}{4CL}(V_E - V_S) = \frac{V_E - V_S}{4f^2CL}$$

siendo $f = 1/T_1$ la frecuencia de la regulación. El valor de C se elige para que la ondulación sea tolerable.

La autoinducción L se elige de forma que nunca se anule la intensidad por la autoinducción (en ese caso la ondulación sería mucho mayor).

Igualando los valores $(\Delta I)_1$ y $(\Delta I)_2$ se obtiene:

$$\frac{V_E - V_S}{L} T_1 = \frac{V_S}{L} T_2 = \frac{V_S}{L}(T - T_1)$$

considerando el período $T = T_1 + T_2$ constante.

Despejando T_1 se obtiene:

$$T_1 = \frac{V_S T}{V_E} \qquad \Delta I = \frac{V_E - V_S}{L}\frac{V_S}{V_E} T$$

Observando la figura 2 y llamando I_{max} al mayor valor de I_s, la condición será:

$$\frac{\Delta I}{2} = \frac{(V_E - V_S)\, V_S}{2LfV_E} < I_{max}$$

es decir:

$$L > \frac{(V_E - V_S)\, V_S}{2fV_E I_{max}}$$

Además no debe saturarse L para $I_L = 2I_{max}$.

12.10.- Reguladores con conmutación en paralelo (Boost-regulator).

La figura 12-13 es un circuito de este tipo:

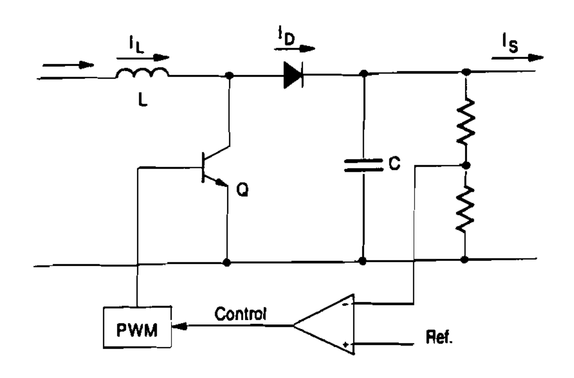


Fig. 12-13.

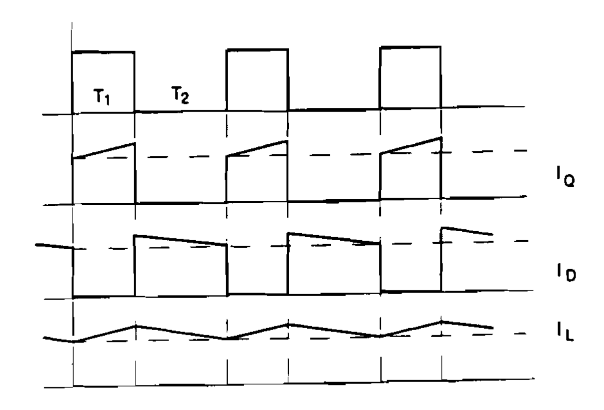


Fig. 12-14.

A diferencia de las fuentes anteriores, en ésta puede obtenerse una tensión de salida V_S mayor que la de entrada V_E, por lo que se les llama "boost" de refuerzo. Se usan poco por el peligro de que crezca demasiado la intensidad por la autoinducción en condiciones no previstas, dando lugar a accidentes destructivos. Su estudio es especialmente interesante para comprender mejor las de transformador, que se verán después.

El funcionamiento puede seguirse a la vista de la Figura 12-14, en que aparecen las formas de onda de las intensidades del circuito.

Durante el tiempo T_1 aumenta la corriente por L, acumulándose energía en la autoinducción. El cambio de corriente es:

$$(\Delta I)_1 = \frac{V_E}{L} T_1$$

Durante el tiempo T_2 la autoinducción invierte su tensión y mantiene la corriente de carga que disminuye hasta

$$(\Delta I_2) = - \frac{V_E - V_S}{L} T_2$$

La tensión negativa de L se suma a la salida para igualar la entrada, con lo que debe ser $V_E < V_S$. En equilibrio se igualan los dos incrementos en valor absoluto, de lo que se deduce el valor de T_1 correcto.

El resto de los cálculos se hace como en el caso anterior.

12.11.- Reguladores con conmutación de retroceso (Flyback regulators) o de transformador.

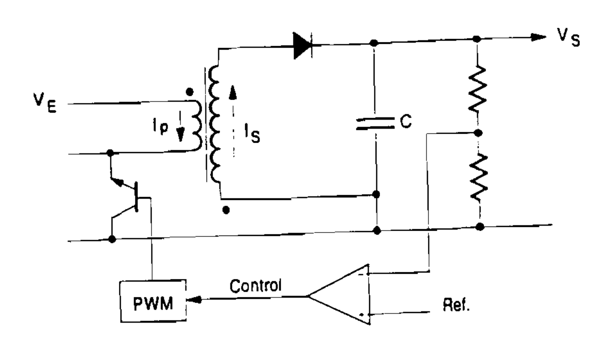


Fig. 12-15

El regulador de retroceso es una variante del anterior. El circuito es el de la figura 12-15. La intensidad producida al cortocircuitar Q acumula energía en la autoinducción del primario del transformador.

Durante el tiempo T_1 crece la intensidad en el primario. La tensión en el primario es la de entrada. La intensidad crece hasta el valor (figura 12-16).

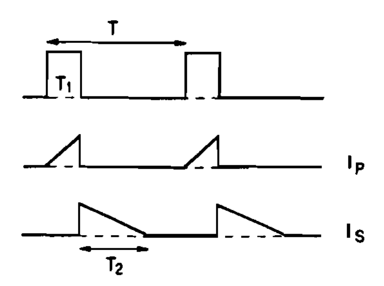


Fig. 12-16.

$$\Delta I = \frac{V_E}{L_P} T_1$$

Durante este tiempo la tensión en el secundario es negativa en el diodo, que no conduce, y su intensidad es nula.

Durante el tiempo T_2 la energía de magnetización del transformador se descarga <u>por el secundario</u>, comenzando con una intensidad de valor $\Delta I/n$ (n relación de espiras) y haciendo conducir al diodo, con lo que la tensión en el secundario será la de salida V_S. La intensidad decrecerá una magnitud

$$\frac{\Delta I}{n} = \frac{V_S}{L_S} T_2$$

donde $L_S = n^2 L_P$. Esta fórmula sólo determina T_2 que debe cumplir $T_1 + T_2 \leq T$, período de regulación.

El condensador se carga a un valor V_S de equilibrio en el que la energía acumulada en T_1 sea igual a la que suministra a la

carga en todo el período. La ondulación se calculará también considerando la cantidad de electricidad que entra en el condensador en cada período.

La presencia del transformador da una gran flexibilidad a estas fuentes, que pueden proyectarse para una amplia gama de tensiones de salida y también para varias salidas (con varios secundarios) aunque la regulación (análoga a la de los casos anteriores, como se ve en la figura 12-15 solo actuará sobre una de las salidas.

12.12.- Onduladores con transistores (Inverters).

La figura 12-17 representa un ondulador (también llamado a veces "inversor" o "convertidor estático de frecuencia") con transistores.

Los transistores actúan como conmutadores que desvían alternativamente la corriente continua por uno u otro lado con lo que el secundario "ve" una forma de onda como la dibujada.

Las formas de onda en las bases y colectores de los transistores de conmutación son las indicadas en la figura 12-18.

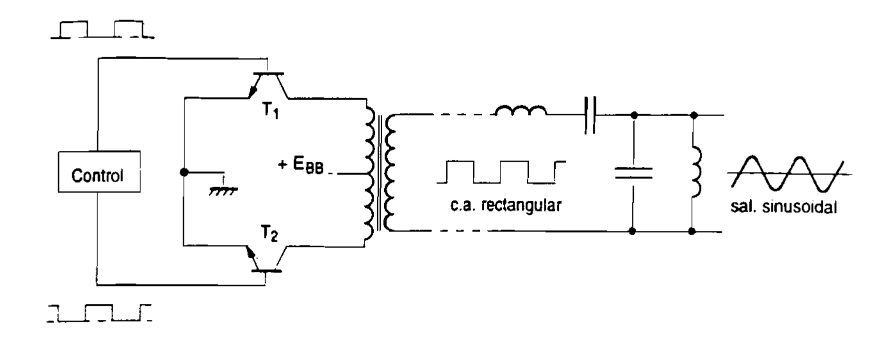


Fig. 12-17.

El control debe suministrar las corrientes de base necesarias para hacer conducir a T_1 y T_2. Cuando empieza a conducir T_1 su colector está a potencial cero y el semiprimario superior está conectado entre $+E_{BB}$ y tierra, con diferencia de potencial $-E_{BB}$. En el colector de T_2 (que queda sin conducir a partir de este momento) aparece la tensión $2E_{BB}$ ya que a la tensión E_{BB} del semiprimario superior se suma otra magnitud E_{BB} <u>inducida</u> en el semiprimario inferior. La diferencia de potencial entre los dos extremos del transformador será $2E_{BB}$.

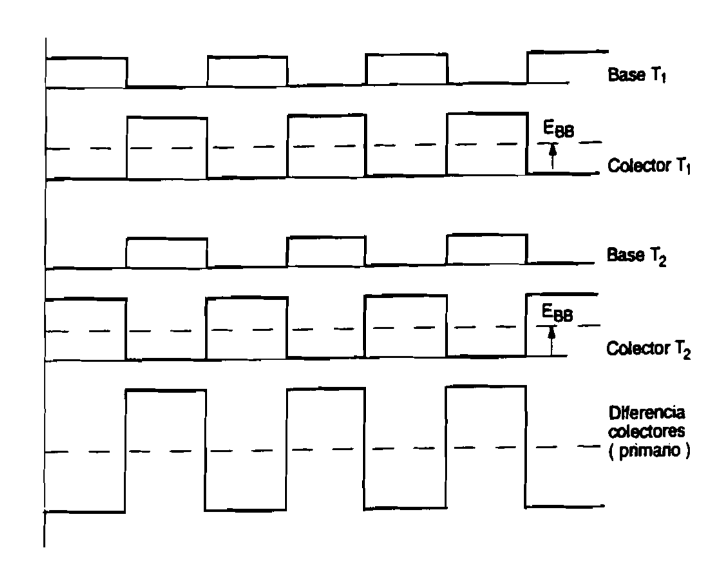


Fig. 12-18.

Al comenzar a conducir T_2 su colector se anula e induce $2E_{BB}$ en el colector de T_1. La diferencia de potencial entre el colector de T_1 y el de T_2, que antes era $-2E_{BB}$, ahora pasa a ser $+E_{BB}$. En cada semiciclo cambia la tensión, de forma que el primario actúa como si estuviese alimentado por una tensión alterna de amplitud $2E_{BB}$. Si la relación de espiras es (1+1):n, la tensión inducida en el secundario será $2nE_{BB}$.

El ondulador puede completarse con un filtro de salida (Figura 7) que seleccione la componente fundamental del desarrollo de Fourier, dando una forma de onda sinusoidal a la salida. La necesidad de este filtrado depende del uso que se va a dar al ondulador. Por ejemplo, si se va a rectificar la salida para obtener otra vez c.c. (¡de valor distinto a la de entrada!) no es necesario filtrar. Por supuesto no se puede filtrar si la salida debe ser de frecuencia "variable" (variable siguiendo la frecuencia de control de las bases).

12.13.- Regulación de fase del ondulador.

Acortando el tiempo de conducción de los transistores de la Figura 12-17 (ver figura 12-19) se puede controlar el valor efectivo de la onda de salida a través de los circuitos de control clásicos, regular la salida.

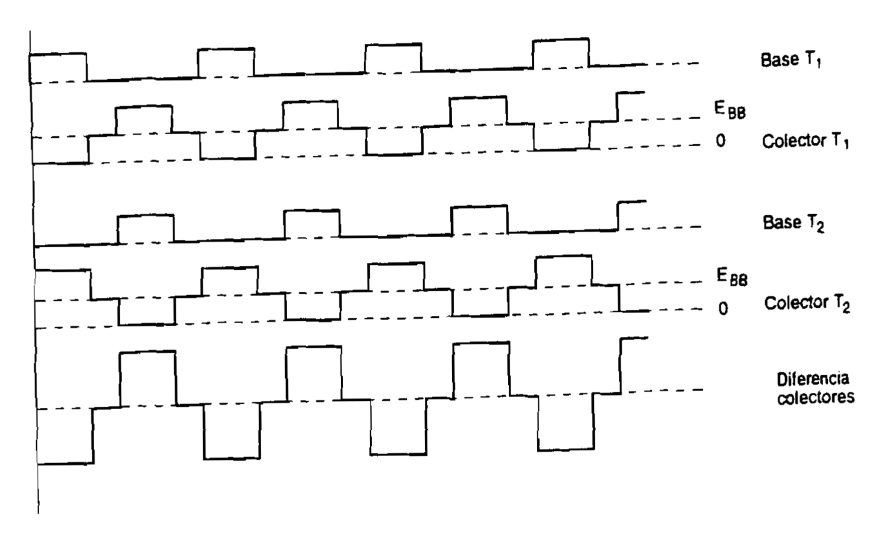


Fig. 12-19.

12.14.- Variante con SCR.

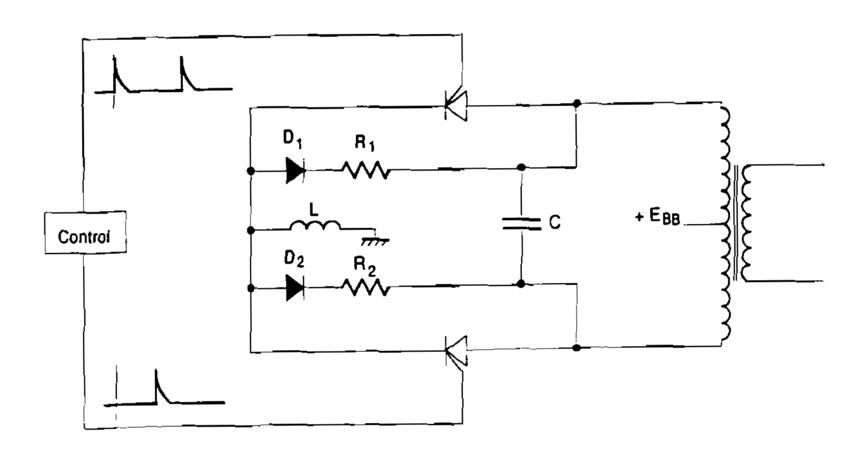


Fig. 12.20.

El circuito de la figura 12-20 es un ondulador en el que la conmutación se hace con tiristores (SCR). Los tiristores tienen la ventaja de que conducen cuando se les aplica un trigger o disparador corto y de muy poca potencia, en lugar de tener que aportar una corriente de base constantemente.

Sin embargo, para que dejen de conducir es necesario aplicar una tensión inversa, lo que se hace con el condensador de conmutación C. Cuando SCR_1 pasa de tener $2E_{BB}$ en su ánodo a cero, el salto $-2E_B$ se aplica a través de C al ánodo de SCR_2 que deja de conducir. El ánodo de SCR_2 pasa rápidamente (en microsegundos) de $-2E_{BB}$ a $+2E_{BB}$ que ya tendrá en el resto del semiciclo. La

descarga de C se hace a través de SCR_2 y, cuando éste se cierra, a través del diodo de conmutación D_2. La autoinducción L sirve para limitar la corriente de la batería a través del transformador que está cortocircuitado durante unos microsegundos.

12.15.- Onduladores resonantes.

Cuando el ondulador es de una frecuencia determinada (por ejemplo 50 Hz ó 400 Hz) puede incluirse el filtro en la carga como se ve en la figura 12.21. En estos casos la salida es directamente sinusoidal. Sin embargo, no se puede cambiar la frecuencia de funcionamiento y las condiciones de carga exterior se hacen críticas. En la figura se ha usado resonancia L_1CL_2, incluyendo L_2 en el secundario del transformador.

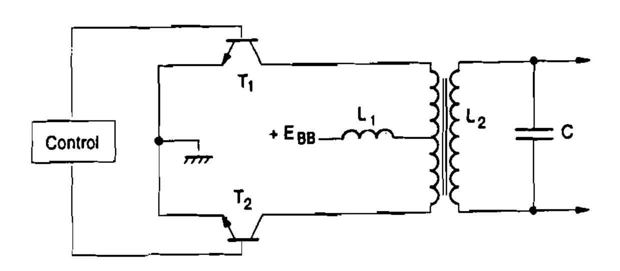


Fig. 12.21.

CAPITULO 13

CALCULO DIGITAL

13.1.- Introducción.

El cálculo digital realiza operaciones lógicas o aritméticas con circuitos electrónicos llamados **digitales o lógicos.** Las señales manejadas por estos circuitos son también digitales, es decir, representan números.

Las señales digitales son funciones del tiempo que sólo toman valores discretos de las variables, o valores asimilados a discretos. Cuando sólo toman dos valores se llaman **binarias.** Por ejemplo, se puede definir una señal binaria con un valor **alto (H)**, que corresponde a una tensión entre un valor umbral y la alimentación (digamos entre 3,5 y 5 voltios) y un valor **bajo (L)**, menor que otro umbral (digamos entre 0 y 1,5 voltios). Los valores intermedios no tienen significado y deben evitarse.

Cuando al valor bajo se le llama 0 y al alto 1, se dice que se está usando **lógica positiva**, que es el caso normal. En el caso inverso, valor bajo 1 y alto 0, se está usando **lógica negativa**, mucho menos frecuente.

También existen funciones digitales que toman más de dos valores, por ejemplo 10 (de 0 a 9), con sus márgenes correspondientes, que definen una **lógica decimal.** En general este tipo de lógica se usa sólo en ordenadores de muy alta velocidad y exige una gran complicación de circuitos. En este libro sólo estudiaremos lógica binaria positiva, que es ampliamente la más usada.

Los componentes que transportan la información digital de un sitio a otro se llaman "buses", palabra inglesa que se ha impuesto y que podría traducirse simplemente por línea digital. La información puede ir **en serie**, es decir, unos números detrás de otros por una sola línea o **en paralelo**, las cifras binarias todas a la vez por un conjunto de líneas.

Los circuitos electrónicos que manejan las señales digitales son circuitos de conmutación, como algunos de los ya estudiados. Los circuitos lógicos pueden dividirse en dos grandes grupos:

a) Circuitos combinacionales o no secuenciales, que efectúan una transformación de la información digital en un momento determinado.

b) Circuitos secuenciales, que efectúan transformaciones de las señales digitales actuales y **anteriores**, usando circuitos de **memoria**. En general tienen un ritmo de actuación, marcado por un **reloj**, circuito que divide el tiempo en intervalos marcando el **tiempo de máquina**, que no tiene porqué coincidir con el tiempo real. Algunos sistemas digitales tienen que intervenir en procesos reales y entonces deben adaptar su ritmo, en general de forma mucho más rápida. Se dice entonces que funcionan en **tiempo real**.

La lógica o herramienta matemática para manejar las señales digitales se llama **álgebra de Boole**. La realización de operaciones complicadas a partir de operaciones elementales constituye la técnica de programación, comúnmente llamada **software**, mientras que los circuitos que la realizan constituyen el **hardware**. La *programación para conseguir manipulaciones* complicadas a partir de las elementales tiene varios niveles. El más sencillo, que simplemente realiza operaciones de correspondencia y secuencia se llama **lenguaje de máquina**. Existen niveles superiores de **lenguaje**, con programación más complicada y más alejada del sencillo lenguaje de máquina. La elaboración de algoritmos de cualquier nivel que realicen funciones determinadas se llama **programación**.

Una combinación de **hardware** o soporte de circuitería adecuado y **software** o herramienta de utilización adecuada ("sistema operativo") constituye un **ordenador** o **computador**, según se utilice la terminología francesa o inglesa, o un **calculador**, con el concepto original de lo que era un sistema de este tipo. Un caso particular de ordenador elemental es el **microprocesador**, que suele ser un sistema adaptado a labores particulares.

En este capítulo empezaremos por estudiar las operaciones elementales de álgebra de Boole con señales binarias y sus circuitos lógicos y secuenciales para pasar a su integración en sistemas digitales. La programación y otras técnicas de software se salen del propósito del libro.

13.2.- Algebra de Boole.

El álgebra de Boole es una teoría deductiva que permite definir dos operaciones fundamentales (y otras derivadas) entre nuestras variables digitales, que, como ya hemos indicado, serán **binarias**, tomando los valores 0 y 1, con lógica positiva (0 es bajo y 1 es alto). Como hemos restringido mucho el campo de aplicación del álgebra de Boole, las operaciones y teoremas fundamentales se justifican fácilmente, sin necesidad de

demostraciones complicadas, como sería el caso si aplicásemos la teoría con todas sus posibilidades.

Las operaciones fundamentales son:

La multiplicación lógica o conjunción **"y"** definida por la siguiente tabla y representada por la **ecuación lógica X = AB**.

A	B	X
0	0	0
0	1	0
1	0	0
1	1	1

La salida es 1 cuando **todas** las entradas son 0 y sólo en ese caso. Esto se puede generalizar al caso de más variables definiendo así la operación X = ABCD... en que el resultado es 1 sólo en el caso de que todas las entradas sean 1.

A la vista de la tabla se ve también que **B es la puerta de A**, ya que si B=0, la salida es X=0 y si B=1, la salida es X=A. También se ven las "propiedades" de que multiplicar por 1 no cambia y multiplicar por 0 da 0.

La suma lógica o disjunción **"o"** (correspondiente al **vel** latino, "o" no exclusivo) definida por la siguiente tabla y representada por **X = A + B :**

A	B	X
0	0	0
0	1	1
1	0	1
1	1	1

La suma es cero cuando todas las entradas son cero, lo que permite generalizar la regla de sumar a más de dos variables. También se comprueba fácilmente que A+0 = A y A+1 = 1. También se puede decir que **B es puerta de A**, en el sentido de que si B=0, X=A, y si B=1, X=1.

Tanto la operación suma como la multiplicación son conmutativas y distributivas, como se deduce fácilmente de las reglas generales.

Una operación de la que ya hemos hablado es la **inversión lógica,** definida por la siguiente tabla y expresada por la ecuación **X =** $\bar{A}$ (se lee equis igual no a):

A	X
0	1
1	0

Estas tres operaciones no son independientes, sino que una están relacionadas por el teorema de Morgan, que se expresa:

$$\overline{AB}=\overline{A}+\overline{B} \quad o\ también$$
$$\overline{A+B}=\overline{A}\ \overline{B}$$

y se comprueba tabulando todas las combinaciones:

A	B	$\overline{A}$	$\overline{B}$	$A+B$	AB	$\overline{A+B}$	$\overline{AB}$
0	0	1	1	0	0	1	1
0	1	1	0	1	0	0	1
1	0	0	1	1	0	0	1
1	1	0	0	1	1	0	0

También se demuestra fácilmente que las operaciones de multiplicación y suma son asociativas, es decir:

$$A\ (B+C) = AB + AC$$

La tabla de demostración sería:

A	B	C	B+C	AB	AC	A(B+C)	AB+AC
0	0	0	0	0	0	0	0
0	0	1	1	0	0	0	0
0	1	0	1	0	0	0	0
0	1	1	1	0	0	0	0
1	0	0	0	0	0	0	0
1	0	1	1	0	1	1	1
1	1	0	1	1	0	1	1
1	1	1	1	1	1	1	1

Del mismo modo se demuestran otras propiedades:

$$(A+B)\ (A+C) = A+BC$$
$$A\bar{A}=0 \quad A+\bar{A}=1$$
$$1+1=1 \quad 1*1=1$$

Se notará la facilidad con que se comprueban las propiedades del álgebra de Boole, debido a que nos referimos siempre a la aplicación a nuestro caso particular. Las demostraciones son más difíciles si se quiere no perder generalidad y las variables no se limitan a representar ceros y unos.

13.3.- Puertas lógicas.

Los circuitos que materializan las operaciones elementales se llaman **puertas**. Ya han sido tratados en el párrafo 11.13. Repetimos aquí la figura 11-29 como fig. 13-1 para comodidad del lector. La realización eléctrica de los diferentes tipos de puertas se vio también en el capítulo de conmutación.

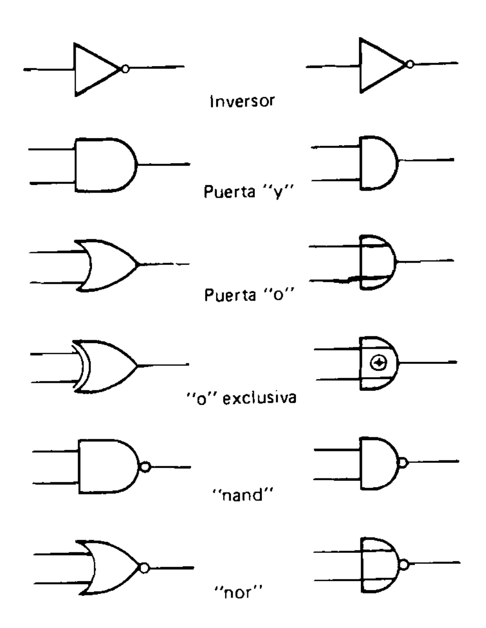


Fig. 13-1.

Las puertas "nor" y "nand" tienen la ventaja de que con un solo tipo de puerta pueden hacerse todas las operaciones, a través del teorema de Morgan o de alguna de las propiedades anteriores. En la figura 13-2 puede verse cómo se hacen las operaciones de inversión, multiplicación y suma lógicas usando sólo puertas "nand" (línea superior) o puertas "nor" (línea inferior).

Inversor   Puerta "y"   Puerta "o"

Fig. 13-2.

13.4.- Funciones de Boole. Teorema fundamental.

Una variable X puede ser función de otras variables A, B, C, ... si se conoce el valor de X para cualquier combinación de las variables. Esta dependencia puede expresarse por una fórmula o por una tabla de correspondencia.

Podemos hallar la tabla que corresponde a una fórmula a través de las operaciones fundamentales definidas. Por ejemplo, calculemos la tabla que corresponde a la fórmula:

$$X=\bar{A}B+A\bar{B}$$

A	B	$\bar{A}$	$\bar{B}$	$\bar{A}B$	$A\bar{B}$	X
0	0	1	1	0	0	0
0	1	1	0	1	0	1
1	0	0	1	0	1	1
1	1	0	0	0	0	0

A la vista de la tabla se ve que se trata del **o exclusivo**, que hemos definido por la tabla.

También es posible sintetizar la fórmula que corresponde a cualquier tabla, utilizando una regla general que se llama **teorema fundamental** y que vamos a ilustrar a la vista del siguiente ejemplo, en el que X es una función de las variables A, B y C. Las tres primeras columnas de la tabla son una enumeración de las combinaciones posibles de tres dígitos 0 y 1. También pueden interpretarse como números sucesivos de 000 a 111 en base 2. El número de combinaciones es $2^3 = 8$, lo mismo que la cantidad de números de 000 a 111 en base 2, ambos inclusive, que también es 2^3. La cuarta columna representa los valores de la función X, por ejemplo, consecuencia de una valoración de cualquier tipo, y que aquí se ha escrito arbitrariamente.

A	B	C	X
0	0	0	1
0	0	1	0
0	1	0	0
0	1	1	0
1	0	0	1
1	0	1	1
1	1	0	0
1	1	1	1

La regla de redacción de la fórmula correspondiente a la tabla es (**teorema fundamental**):

La fórmula se compone de la suma de tantos términos llamados "minitérminos" como unos haya en la columna de la X, cuatro en el ejemplo. Cada término corresponde a una fila cuyo resultado X sea 1.

Cada término se escribe como producto de todas las variables (tres en el ejemplo) que se escriben directamente cuando su valor en esa fila sea 1 o negadas cuando el valor en esa fila sea 0.

Así, la fórmula que corresponde a la tabla citada será:

$$X=\overline{A}\ \overline{B}\ \overline{C}+A\ \overline{B}\ \overline{C}+A\ \overline{B}\ C+A\ B\ C$$

Esta regla se justifica fácilmente leyendo el polinomio como función "o" de la siguiente forma: X es 1 cuando coinciden los valores 1 de los inversos de A, B, C, es decir coinciden A,B y C en cero, **o** cuando coinciden los valores 1 del inverso de A (A=0), 1 del inverso de B (B=0) y 1 de C (C=1), etc... como se ve en la tabla.

La fórmula obtenida con esta regla no es única, ni tampoco la más sencilla. Se puede simplificar teniendo en cuenta las propiedades de las operaciones de Boole. Así por ejemplo, los dos primeros términos y los dos últimos pueden simplificarse aplicando la propiedad distributiva y la propiedad de que la suma de una variable y su inversa es siempre 1 :

$$X=\overline{A}\ \overline{B}\ \overline{C}+A\ \overline{B}\ \overline{C}+A\ \overline{B}\ C+ABC=(\overline{A}+A)\overline{B}\ \overline{C}+AC(\overline{B}+B)=\overline{B}\ \overline{C}+AC$$

La fórmula que corresponde a una tabla se llama también **ecuación de aplicación.**

13.5.- Realización de correspondencias lógicas con puertas.

Una tabla de corresponedencia lógica da lugar a una fórmula compuesta de minitérminos. Cada minitérmino es una multiplicación o un "y" lógico y la suma de los términos puede realizarse con un "o" lógico. La figura 13-3 es la realización con puertas lógicas de la fórmula del párrafo anterior, es decir, se materializa la función:

$$X=\overline{B}\,\overline{C}+AC$$

usando circuitos "no", "y" y "o".

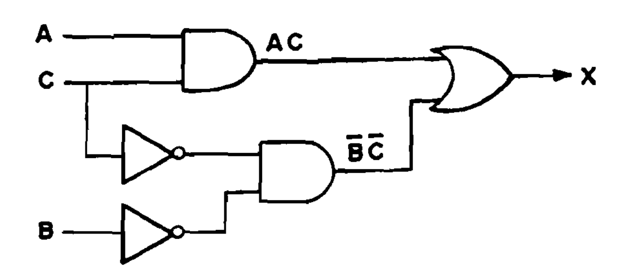


Fig. 13-3.

En la figura se ha supuesto que se dispone de las variables A, B y C, pero no de sus inversas. Si se entra también con los inversos (seis entradas), no es necesario poner inversores.

Como ya hemos indicado se pueden realizar todas las operaciones lógicas con un solo tipo de puertas, si éstas son "nor" o "nand".

13.6.- Matriz de diodos.

Un modo práctico de obtener una o varias funciones de Boole de varias variables es la **matriz de diodos** representada en la figura 13-4. Las líneas horizontales son las entradas, están conectadas a las variables de entrada, directas e inversas. Las líneas verticales son los minitérminos, unidos a las variables de entrada correspondientes mediante diodos que materializan las puertas "y". Por ejemplo, en la figura se han materializado los términos $A\overline{B}\,C, AB$ y $\overline{A}\,C$. .

Las salidas verticales se unen luego con diodos para formar las sumas que sean necesarias para obtener polinomios de salida. Por ejemplo, en la figura se han "cableado" las salidas $A\overline{B}\,C$ y $X=AB+\overline{A}\,C$.

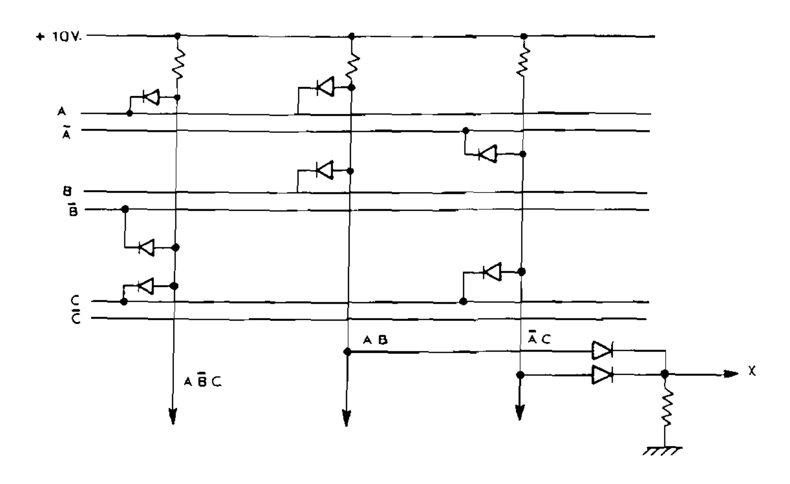


Fig. 13-4.

Una matriz de este tipo puede ser programada en fábrica (por encargo) o traer **todos** los diodos entre líneas horizontales y verticales. Cada diodo lleva en serie un pequeño fusible que se puede quemar desde el exterior, aplicando una fuerte corriente entre la salida y entrada que corresponda, quitando así los diodos no necesarios y deja los que correspondan a la ecuación de aplicación. Una matriz de este tipo se llama **dispositivo lógico programable, PLA** (programable logic array). También suelen llevar incluidos los inversores de entrada, con lo que sólo es necesario dar entradas directas.

13.7.- Circuitos flip-flop.

El circuito flip-flop ha sido ya estudiado en el capítulo de conmutación. Ya allí se indicó que es un circuito ideal para **memoria**, ya que conserva el **estado** en que se dejó hasta que se cambie desde el exterior. Naturalmente, el circuito se "olvida" del estado en que estaba si falta la alimentación. Por esta razón se dice que es una **memoria volátil**.

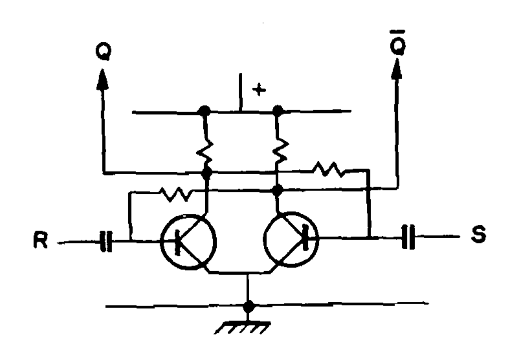


Fig. 13-5.

Existen variaciones de circuitos flip-flop, según la forma

de conexión de las entradas, que hacen que la función lógica sea diferente, y que describiremos a continuación.

En la figura 13-5 reproducimos el flip flop del capítulo 11, figura 11-17, conectando las entradas mediante condensadores que dejan pasar los impulsos, o, más bien, convierten en impulsos los frentes de onda. En esta figura se han representado las **entradas S, set** (poner en 1), **R, set** (poner en 0, borrar) y las **salidas**, **Q** que refleja el estado 1 o 0 del flip-flop y $\overline{Q}$ que refleja el estado inverso. Las entradas son impulsos positivos, que pueden proceder de diferenciar puertas (borde de entrada). También se pueden usar impulsos negativos, con lo que R y S intercambian sus funciones. Si se quiere evitar que "pasen" impulsos de signo contrario, se agrega un diodo en paralelo con la entrada, que cortocircuita los impulsos de signo a favor del diodo.

Este tipo de flip-flop se llama **RS**. está compuesto por dos inversores (los dos transistores), por lo que se puede representar con dos puertas lógicas **NAND** usadas como inversores y complementadas con las entradas exteriores, como se ve en la figura 13-6. También se pueden usar los símbolos de negación cuando se usan impulsos negativos como se ve en la misma figura. En ella se incluyen los símbolos del flip-flop directo (a la izquierda) o con las entradas invertidas (a la derecha).

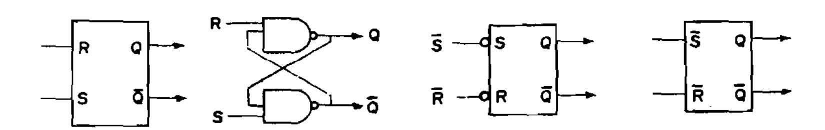


Fig. 13-6.

Flip-flop J-K.

En el flip flop RS no está claro el estado cuando se dan entradas R y S a la vez (depende de la relación de tiempo de las entradas). Para hacer que, de forma bien definida, cambie de estado cuando se dan las dos entradas, se usa una **lógica de entrada** con una **ecuación de aplicación**, que corresponde a la tabla de correspondencia que dice lo que deben ser R y S en función de las entradas exteriores, que ahora se denominan J y K y **del estado anterior Q**, que se realimenta de la salida.

La tabla de correspondencia es la que se da a continuación. Como se ve en la tabla, se han escrito las 8 combinaciones posibles de entradas J y K y del estado anterior Q, escribiendo a continuación cual debe ser la salida Q en cada caso. A la vista del estado anterior y del siguiente , se determina si hay que hacer S=1 y R=0, para cambiar de 0 a 1, o bien S=0 y R=1, para cambiar de 1 a 0, o bien, si se trata de "dejarlo" en 0 , S=0 y R puede ser 0 o 1, indiferente, lo que se expresa con una **x**, o bien dejarlo en 1, con R=0 y S = x, **indiferente**.

Entrada J	Entrada K	Estado anterior	Estado sig.	Valor de S	Valor de R
0	0	0	0	0	x
0	0	1	1	x	0
0	1	0	0	0	x
0	1	1	0	0	1
1	0	0	1	1	0
1	0	1	1	x	0
1	1	0	1	1	0
1	1	1	0	0	1

Una vez escrita la tabla, pueden obtenerse las **ecuaciones de aplicación**, que dan R y S en función de J,K y Q (estado anterior). El valor de x puede tomarse 0 o 1 según convenga para simplificar las ecuaciones. En este caso (en que x está en la salida) conviene siempre 0. Las ecuaciones de Boole las simplificamos de la forma conocida.

$$S=J\overline{K}\,\overline{Q}+JK\overline{Q}=J\overline{Q}(\overline{K}+K)=J\overline{Q}$$
$$R=\overline{J}\,KQ+JKQ=KQ(\overline{J}+J)=KQ$$

El circuito lógico que materializa estas ecuaciones es el de la figura 13-7:

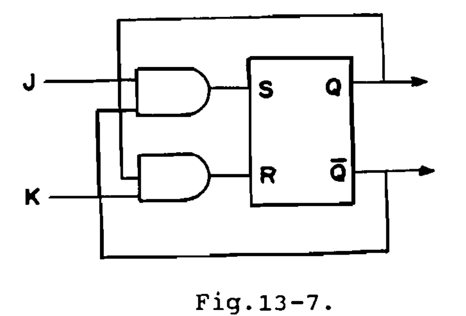


Fig.13-7.

Flip- flop RS síncrono.

Es un circuito RS en el que las entradas están "pasadas" por dos puertas controladas por una tercera entrada llamada **reloj (clock)**. El circuito obedece las órdenes de entrada sólo cuando, simultáneamente, llega un impulso positivo del reloj. La figura 13-8 es un ejemplo en el que se ve el diagrama lógico y el símbolo. Como en casos anteriores, las entradas pueden estar "negadas", lo que se expresa con un circulito o con el símbolo

de negación sobre la entrada.

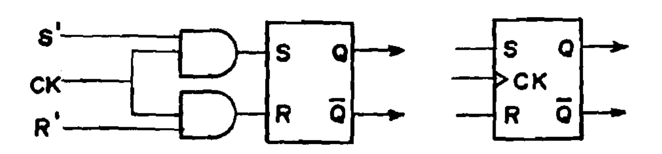


Fig. 13-8.

Flip- flop tipo D.

Es un circuito **seguidor digital o cerrojo (lock),** que toma el estado marcado por la entrada cuando llega el impulso del reloj. La figura 13-9 representa el diagrama lógico basado en un RS.

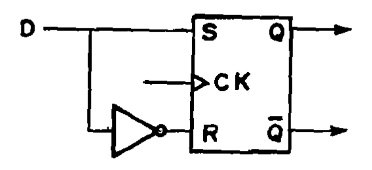


Fig. 13-9.

Flip- flop tipo T.

Es un circuito que cambia de estado cuando recibe un impulso denominado **trigger (disparador).** Puede ser un JK que recibe J y K cuando llega el reloj, como se ve en la figura 13-10. También puede incluirse un inversor y funcionar con entrada negada.

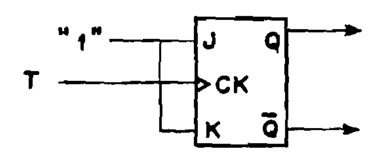


Fig. 13-10.

Flip- flop JK síncrono.

Es un circuito JK en que las entradas solo actúan cuando llega el impulso del reloj, de forma análoga a como se ha descrito en el caso del RS, y como se ha usado en la figura 13-10 como parte del circuito T.

Tanto en los JK como en otros, a veces se dejan accesibles las entradas RS con puertas "o" de forma que pueda actuarse sobre ellas independientemente de otras entradas. En algunos, se denominan estas entradas **PR (preset)**, por set, y **CLR (clear)**, por reset. La figura 13-11

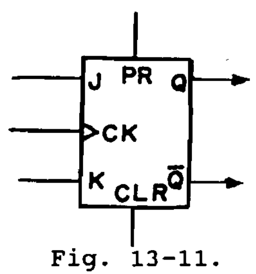


Fig. 13-11.

13.8.- Contadores.

La operación de contar, en base 2, se puede descomponer en dos operaciones lógicas:

Cada vez que se recibe un impulso, se cambia la cifra de unidades. Cada vez que una cifra pasa de 1 a 0 se cambia la cifra de orden superior (arrastre). La figura 13-12 es un ejemplo de contador elemental, basado en flip-flops de tipo T. La negación de entrada T indica que cuenta impulsos (o flancos) negativos. El arrastre se toma de la salida Q de cada cifra, que da un impulso negativo al pasar de 1 a 0 y se lo pasa a la cifra superior (los flip-flops de este contador corresponden a orden superior hacia la derecha, al revés de la escritura normal de un número de varias cifras).

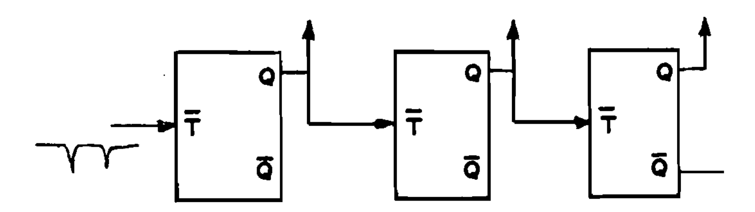


Fig. 13-12.

La operación de arrastre, típica de los contadores, puede hacerse con un algoritmo más complicado, contando en un orden arbitrario o en dirección contraria. Esto se materializa con **ecuaciones de aplicación** en cada flip-flop, que sustituyen a la simple conexión de la figura 13-12.

La figura 13-13 representa un contador decimal de base 2, llamado **década**. La función T se realiza aquí con la entrada de reloj (invertido) y conexión J y K a 1. Además, se incluye un circuito lógico que pone el contador en cero cuando se llega a **10 = 8 + 2**, cambiando a cero a través de la entrada CLR. También se puede poner en cero accediendo a la "barra" de CLR desde el exterior, a través de una puerta "o", que pone en 0 cuando se

ordena desde el exterior o cuando se ha llegado a 10 (se dice que se ha llenado el contador).

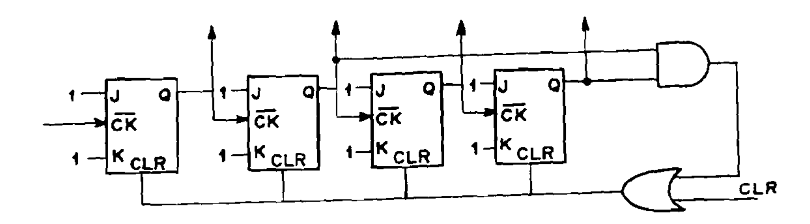


Fig. 13-13.

13.9.- Registros y circuladores.

El flip-flop constituye un tipo de memoria "corto plazo" muy adecuado para anotar números que constituyen los datos para la unidad aritmética y para otras operaciones o rutinas.

Un conjunto de flip-flops preparado para anotar números se llama un **registro**. La información numérica se pasa de unos registros a otros con **puertas** que controlan el tráfico de la información. La figura 13-14 es un esquema de paso de la información del registro (de tres cifras) superior al inferior, previamente borrado (a través de un "CLR"). Si no estuviese borrado habría que pasar también la información de las salidas invertidas a las entradas R, o utilizar inversores para obtener las salidas invertidas. El paso se hace con una sola orden de control, que abre las puertas correspondientes.

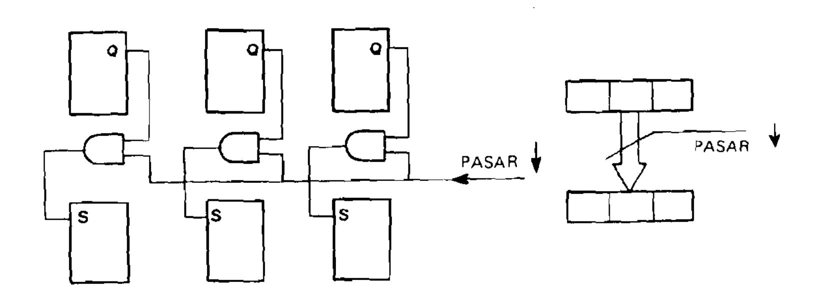


Fig. 13-14.

A la derecha de la figura se ve el símbolo de los registros y de los "canales de paso" en paralelo, que constituyen un **bus** de tránsito de información **en paralelo**.

Muchas operaciones se hacen **en serie**, una cifra después de otra, como se hacen las operaciones aritméticas tradicionales. Por ejemplo, para sumar en base 10 dos números de varias cifras, primero se suman las unidades, se escribe el arrastre y, a continuación se suman las decenas, y así sucesivamente.

En cálculo digital, se empieza igual, es decir, se suman las unidades y se escribe el arrastre. Pero, en lugar de avanzar para la suma siguiente, se mueven los datos, desplazándolos de forma que las decenas se coloquen donde antes estaban las unidades, y así se utiliza un circuito que siempre suma sobre la misma posición.

Para desplazar las cifras sucesivamente, se utilizan los **circuladores.** La figura 13-15 representa un **circulador.** Cuando se recibe una orden de circular, a cada impulso del reloj cada flip-flop toma el estado del flip flop anterior, ya que se trata de circuitos D que siguen la instrucción Q. Así se desplazan los dígitos un lugar hacia la derecha y el último pasa a ser el primero. Cuando han transcurrido cuatro impulsos del reloj, todos han vuelto a su posición inicial. Si se desea dar una vuelta, se dará una orden de circular que comprenda cuatro impulsos de reloj.

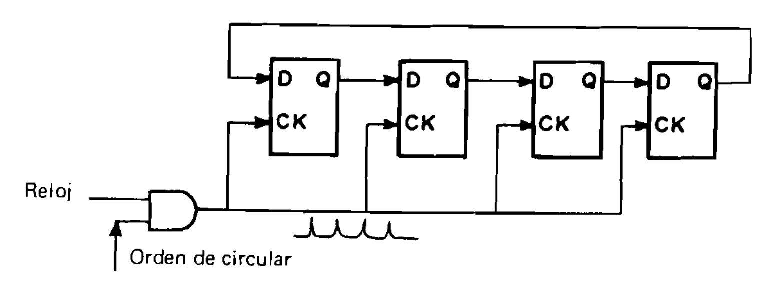


Fig. 13-15.

13.10.- Sumadores.

Como ejemplo de aplicación de los circuitos anteriores, vamos a ver cómo se suman "en serie" dos números de cuatro cifras en base 2, escritos en los registros a_4 a_3 a_2 a_1 y b_4 b_3 b_2 b_1. (fig.13-16). A continuación escribimos la tabla de sumar:

m_{i-1}	a_i	b_i	Suma c_i	Arr. m_i
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

En las tres primeras columnas escribimos las 8 combinaciones posibles de tres números de un dígito (se toman tres para tener en cuenta el arrastre). En esta tabla, se suman a_i, b_i y el arrastre m_{i-1}. La cuarta columna es la suma de las tres anteriores c_i y la quinta columna es el nuevo arrastre m_i.

A partir de esta tabla pueden escribirse las ecuaciones de Boole del sumador, que son:

$$c_i=\overline{m_{i-1}}\,\overline{a_i}b_i+\overline{m_{i-1}}a_i\overline{b_i}+m_{i-1}\overline{a_i}\,\overline{b_i}+m_{i-1}a_ib_i$$
$$m_i=\overline{m_{i-1}}a_ib_i+m_{i-1}\overline{a_i}b_i+m_{i-1}a_i\overline{b_i}+m_{i-1}a_ib_i$$

La realización eléctrica de estas dos ecuaciones (que pueden simplificarse) da lugar a un **sumador** de tres entradas y dos salidas, que denominaremos Σ.

La disposición completa del sistema puede apreciarse en la figura 13-16

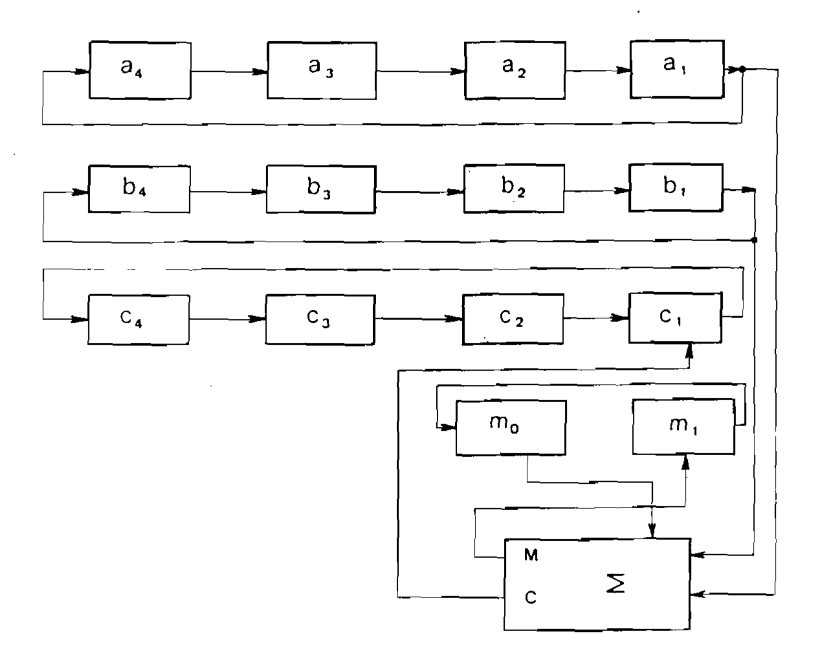


Fig. 13-16.

Las cuatro líneas superiores son cuatro circuladores. Inicialmente, se escriben los dos números de cuatro dígitos en los circuladores a_i, b_i y ceros en el m_0, m_1. Al dar la orden de sumar (no indicada en la figura) funciona el sumador, que efectúa la suma de a_1, a_2 y $m_0 = 0$ y escribe el resultado en c_1 y el arrastre en m_1. A continuación se da la orden de circular, los

valores de a_2 y b_2 pasan al primer lugar (en lugar de movernos nosotros como operador hacia la izquierda, como se hace en operaciones manuales, se mueven los números hacia la derecha para conectarse a la entrada del operador-sumador). El valor de arrastre escrito en m_1 pasa a m_0, para ser sumado también. El sumador efectúa una nueva suma y escribe el nuevo resultado c_2 en el lugar primero, el de c_1 y el nuevo arrastre en m_0. La operación se repite otras dos veces más (cuatro en total). Después de cuatro circulaciones los números a_i y b_i están de nuevo en su sitio, en el registro c_i está escrito el resultado y la quinta cifra de resultado, si la hay, está en m_1. Si no se quiere que los resultados tengan más cifras que los datos, la aparición de la quinta cifra se interpretará como un "overflow", sobrecarga.

Los impulsos de control deben estar cuidadosamente secuenciados para evitar que se hagan dos cosas a la vez. Por ejemplo, en los ordenadores es muy frecuente que el reloj dé dos "fases" de impulsos alternativos: la primera fase hace el tráfico de información, como circulación etc... y la segunda fase las operaciones. Con eso se evita que se desplace un dígito mientras se está operando, con la consiguiente confusión. También debe estar muy bien estudiada la **secuencia** de operación para evitar el mismo tipo de confusiones (es decir, primero escribir los datos, después hacer las operaciones y finalmente extraer los resultados para su presentación).

El sumador descrito lo es **en serie** porque hace la operación dígito a dígito. Pero también se puede hacer la suma **en paralelo**, toda a la vez. Para ello, en el ejemplo anterior, se escribiría una tabla de 2^8 = 256 combinaciones posibles de a_i y b_i y se obtendrían 5 funciones de Boole que darían los cuatro valores de c_i y el último arrastre. La materialización de este sumador es naturalmente mucho más complicada que la del sumador serie, pero la rapidez es mucho mayor, ya que hace toda la operación de una década con un solo impulso de reloj.

13.11.- Memorias.

Frecuentemente es necesario anotar grandes cantidades de dígitos para su utilización posterior. Una memoria es un dispositivo capaz de guardar esa información. Un registro es también una memoria, pero su capacidad es reducida, de uno o varios "bytes". Un byte es un conjunto de un número de bits determinado, típico del ordenador o sistema, que puede ser 8, 16 o 32 bits según los casos. La denominación de memoria se reserva para dispositivos de muchos miles de bytes o "kilobytes".

Las memorias en que se puede escribir un byte en un sitio o **dirección** determinada sin pasar por los demás, es decir, que se entra en el dispositivo con la dirección y el dispositivo da el contenido de esa dirección se llaman **RAM**, de acceso aleatorio **(random access memory)**, denominación no muy afortunada, pero que expresa que no hay que seguir un orden determinado para "preguntar" y que implica que son muy rápidas, precisamente por esa rapidez de acceso. En general, en este tipo de memorias se

puede escribir, borrar o leer con la misma rapidez. Las memorias RAM pueden ser **volátiles,** si se pierde la información cuando falla la alimentación o **no volátiles** si se conserva la información independientemente de la alimentación.

Existen otras memorias que solo se pueden leer, aunque con la misma rapidez. Aunque también son de acceso aleatorio se llaman **ROM, (read only memory)**. En ellas se almacenan datos fijos que no es necesario renovar.

También existen otras memorias que no son de acceso rápido o "aleatorio" sino que es necesario acceder a cada posición pasando por otras, por ejemplo, discos magnéticos u ópticos. Se llaman secuanciales y se caracterizan por tener una mayor densidad de información, es decir mayor capacidad de almacenar información en el mismo volumen.

13.12.- Memorias estáticas de semiconductores.

El elemento básico es el **flip-flop** que ya hemos estudiado. El flip-flop puede ser de transistores (**TTL**) o de **MOSFET**. La figura 13-17 representa un flip-flop de transistores en el que se han incluido los transistores de control que dan acceso al flip-flop. La figura 13-18 representa la agrupación de elementos en forma de matriz de 128 (2^8) bytes de 8 bits, es decir, un total de 1.024 bits. Cada una de las 128 filas puede guardar la información de 8 bits, en 8 flip-flops.

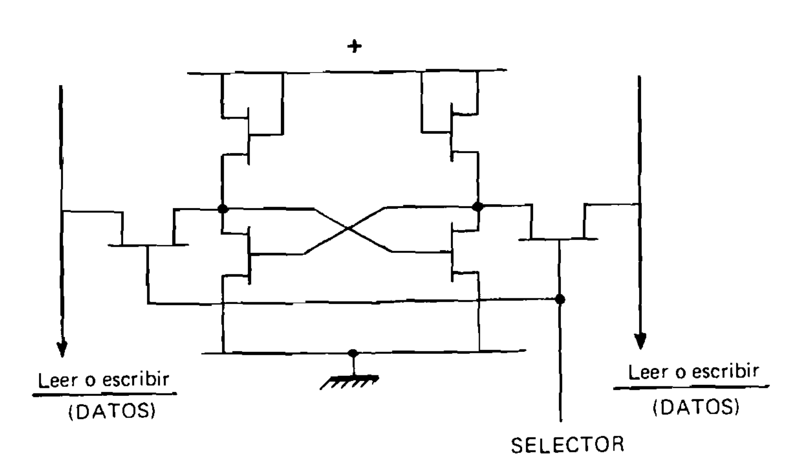
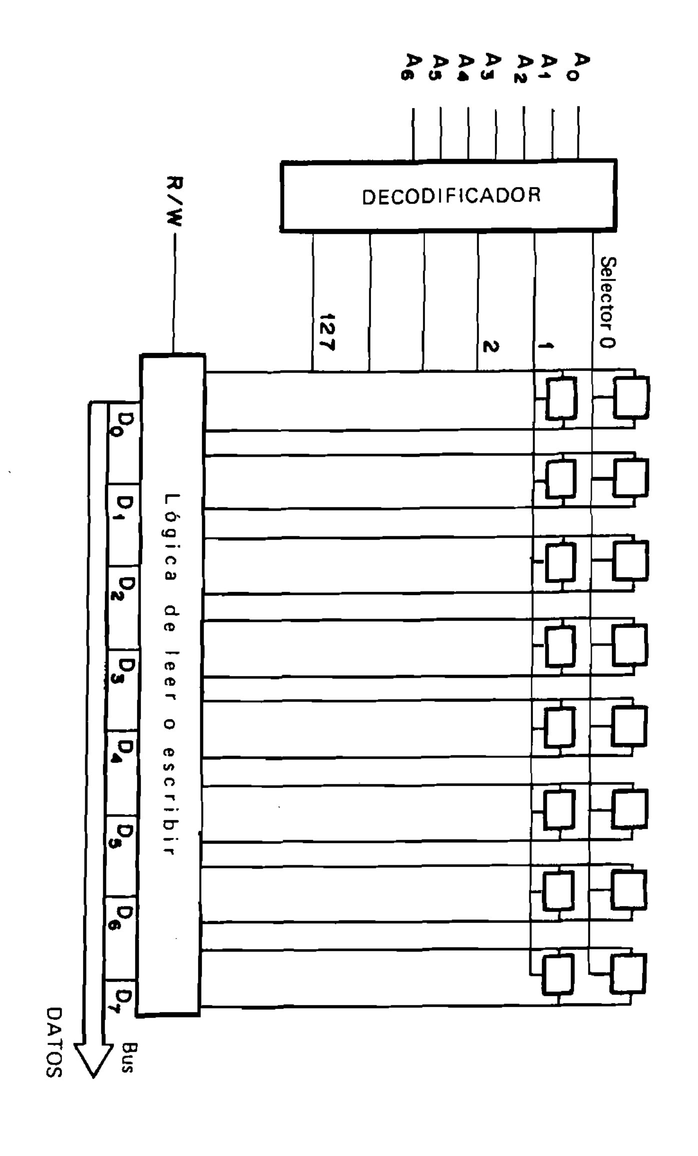


Fig. 13-17.

La decisión de qué fila se lee o escribe la toma un **decodificador** de 7 entradas binarias que obtiene con circuitos lógicos o con una matriz de diodos las 128 (de 0 a 127) direcciones de los bytes de la memoria. Cuando se da nivel alto a los controles de una fila se puede acceder a los flip-flop desde el **bus** de 8 líneas dobles de **datos**, que se leen o escriben en los 8 flip-flops de la línea seleccionada. Las 16 líneas de datos manejan la información de los 8 datos D_1 D_2 D_3 D_4 D_5 D_6 D_7 D_8 y sus inversos.



A_0 A_1 A_2 A_3 A_4 A_5 A_6
DECODIFICADOR
R/W
Selector 0
1
2
127
Lógica de leer o escribir
D_0 D_1 D_2 D_3 D_4 D_5 D_6 D_7
Bus
DATOS

Fig. 13-18.

El bus de datos tiene 8 líneas dobles que pueden actuar de transmisoras o receptoras a través de puertas que se abren en una u otra dirección y que se llaman **buffer** (separadores). El sentido del tráfico de la información se controla con una orden llamada **R/W (read/write)** de la que el nivel alto corresponderá a leer (R) y el bajo a escribir (W).

Cada flip-flop queda conectado a una de las líneas de datos por sus dos entradas (fig. 13-17), tanto para escribir como para leer.

Cuando se trata de escribir, el flip-flop actúa como un R/S convertido en D, ya que recibe la información directa y negada del dato correspondiente.

Cuando se trata de leer, el flip-flop envía su estado, directo e invertido a la línea de datos, a través del buffer.

13.13.- Memorias dinámicas de semiconductores.

Están constituidas por condensadores que se cargan con una tensión determinada para escribir un 1 y permanecen largo tiempo cargados. Cuando el condensador se descarga se le atribuye un 0. La carga y descarga se hace a través de un MOSFET de "capacitación" que se selecciona mediante el algoritmo de "direccionamiento" descrito en el párrafo anterior (fig. 13-19).

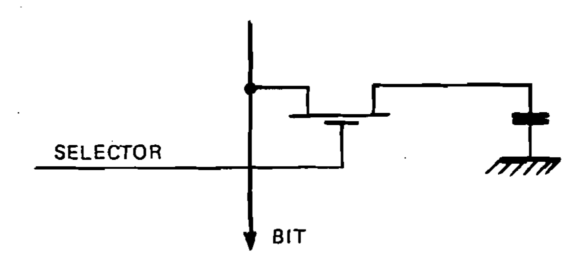


Fig. 13-19.

Estas memorias son más sencillas que las de flip-flops, se pueden miniaturizar más (unas cuatro veces) y tienen menos consumo. El inconveniente de estas memorias es que los condensadores pierden su carga lentamente y es necesario disponer de un costoso sistema de **refresco** que lea todos los condensadores y los recargue nuevamente de forma periódica.

13.14.- Memorias ROM.

Se llaman memorias **ROM (Read Only Memory)** a dispositivos de acceso aleatorio (es decir que también son RAM , aunque no se les denomine así) en los que se puede leer una gran cantidad de información programada por el fabricante o preescrita eléctricamente por el usuario antes de utilizarlos en su sistema.

En general puede decirse que las memorias ROM responden a una

entrada llamada **dirección** con una salida de datos, que es la información contenida en esa dirección. Más que una memoria, se trata pues de una **correspondencia lógica** de la salida con la entrada, del tipo no secuencial, análoga a la de las matrices de diodos. Sin embargo se presentan en forma de memorias y así las seguiremos denominando.

Memorias ROM.

Se trata de una matriz de diodos, construidos en el IC (circuito integrado) por el fabricante, con un sistema de máscara en las fases de difusión. Si, por ejemplo, se trata de una memoria de 128x8 bits, como la representada en la figura 13-20, la dirección de entrada (de 7 bits, 128 números) se decodifica y alimenta una de las 128 líneas horizontales de la figura. Supongamos que se alimenta la línea número 1. Si el número que se quiere extraer de esa posición 1 es el 10110100, el fabricante conecta diodos en los "1" y no lo hace en los "0", con lo cual las líneas verticales primera, tercera, cuarta y sexta reciben tensión alta y la pasan a las ocho líneas que constituyen el bus de salida.

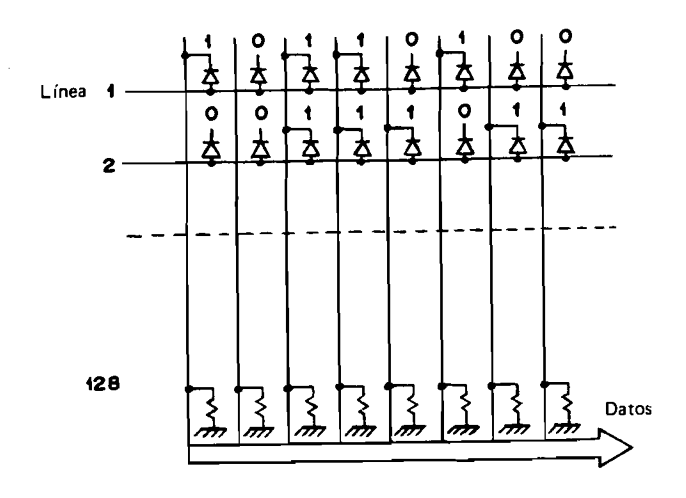


Fig. 13-20.

En realidad se trata de un PLA (programable logic array) como los que ya hemos visto.

Memorias PROM.

Se trata de memorias ROM como las anteriores, pero programables por el usuario. Las conexiones de los diodos se hacen mediante microfusibles que se abren cuando el usuario da una tensión suficiente entre las entradas y salidas, escribiendo así un 0. La programación solo se puede hacer una vez, ya que los fusibles no pueden recuperarse.

Memorias EAROM.

Se trata de una matriz en que las conexiones están hechas con MOSFET con puertas aisladas que pueden **cargarse** para que el MOSFET conduzca, inyectando las cargas a través del dieléctrico o descargarse iluminando fuertemente con luz ultravioleta.

13.15.- Codificación.

En cálculo digital es frecuente usar diferentes métodos de representación numérica o **códigos**. Aunque existen códigos con cifras decimales y otras, aquí solamente trataremos la codificación binaria, mucho más habitual.

La forma más simple de codificar un número en binario es la codificación binaria directa, en que los números se representan en base 2. Pero hay otras variantes que veremos a continuación. En todas ellas, se han desarrollado **algoritmos** que permiten hacer operaciones y manipulaciones sin necesidad de pasar por códigos binarios directos. El desarrollo de este tipo de algoritmos constituye uno de los objetivos de una ciencia especial, la **informática**.

Código decimal binario.

Es una representación decimal en que cada cifra en base 10 se expresa en binario. Se llama **BCD** (binary code decimal). En realidad es una etapa de representación que facilita la conversión a decimal. Cada década es un registro de 4 flip-flop que, en realidad, tiene capacidad para 16 dígitos, utilizándose sólo para 10.

Código hexadecimal.

Las cifras están en base 16. Con ello se aprovechan totalmente los registros de cuatro flip-flops. Los dígitos en base 16 se llaman 0,1,2,...,9,A,B,C,D,E y F. Por ejemplo, B en base 16 vale 11 en base 10. Tiene la ventaja de que expresando los dígitos hexadecimales en base 2 (cuatro por cada uno), se obtiene directamente la codificación en base 2 . Es decir, la representación de un número de n dígitos de base 16 se pasa a la representación de 4n dígitos en base 2 expresando cada dígito por separado.

Complemento a 2.

Es la representación binaria de 2_n -X para el número X de menos de n dígitos en base 2. Se utiliza directamente como representación de los números negativos. El complemento a 2 de X representa a -X para restar. La obtención del complemento a 2 se hace cambiando los 0 por 1 y los 1 por 0 y sumando 1, como es fácil demostrar.

Complemento a 1.

Es una variante del anterior. Se obtiene cambiando los 0 por 1 y los 1 por 0 en X, por lo que vale 2^n - X - 1.

Código Gray.

Está basado en que la representación de cada número difiere del anterior en un solo dígito. La tabla de correspondencia de 0 a 9 es:

0	0
1	1
2	11
3	10
4	110
5	111
6	101
7	100
8	1100
9	1101

La ventaja de este tipo de codificación es que al no cambiar varios dígitos a la vez, las transiciones en los sensores de posición (por ejemplo, angulares) son menos bruscas.

Códigos redundantes.

Son códigos en que pueden darse reglas para detectar errores aleatorios que no obedecen a la codificación. Por ejemplo, puede introducirse un **bit de paridad,** de forma que el número de unos sea siempre par. Si aparecen errores aleatorios, aparecerán también números que no cumplen la condición de paridad, y podrá detectarse su presencia.

13.16,. Ordenadores.

Un **ordenador o computador** es una máquina capaz de efectuar operaciones programadas con arreglo a ciertas reglas, a partir de las entradas y de los valores almacenados en su memoria interna.

El modelo teórico de **ordenador secuencial** (es secuencial en cuanto tiene en cuenta valores anteriores) es el llamado modelo de **"Hoffman- Moore",** representado en la figura 13-21. Si representamos por x_{in}, y_{in}, z_{in} el valor de las variables x_i, y_i, z_i en el momento n-simo (tiempo interno del ordenador o **tiempo de máquina**), el modelo responde a las ecuaciones:

$$y_i^n = g_n(x_1, x_2, \ldots x_p, y_1, y_2, \ldots y_q, z_1, z_2, \ldots z_r)^{n-1}$$
$$z_i^n = f_n(x_1, x_2, \ldots x_p, y_1, y_2, \ldots y_q)^n$$

Es decir, se escribe en la memoria una determinada función de las variables de entrada, de las de la propia memoria y de las de salida **del momento anterior** (se hace en el momento anterior) y se obtiene en cada momento una salida función de las entradas y de la memoria. Si se trata de una máquina binaria, todas las variables pueden tomar los estados 0 y 1 (fig.13-21).

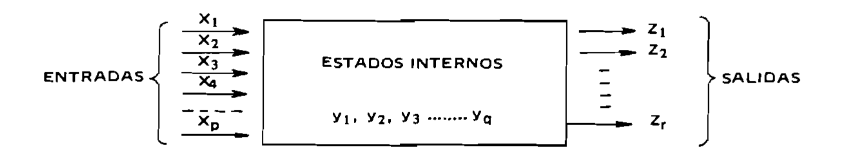


Fig.13-21.

El ordenador se descompone en grandes bloques funcionales, representados en la figura 13-22.

El dispositivo de entrada es un traductor de los datos e instrucciones de entrada en código eléctrico. Puede ser un teclado, un disco o cinta magnética, un disco óptico, o una conexión con otro ordenador.

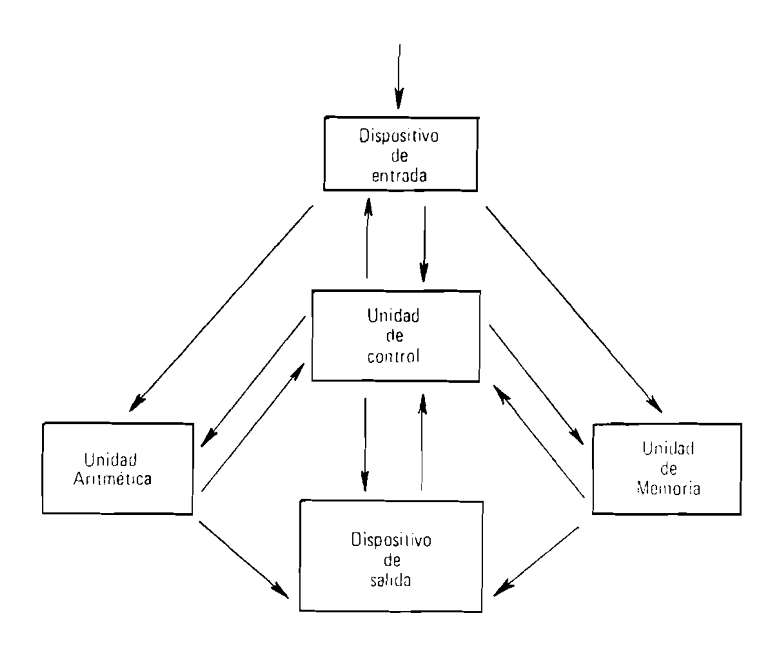


Fig. 13-22.

La unidad de control fija el ritmo de trabajo con un **reloj** y coordina todo el tráfico de información por canales o **buses**, enviando la información de una unidad a otra, detectando las instrucciones y separándolas de los datos, etc...

La unidad de memoria (que puede estar subdividida en varias) mantiene la información clasificada en instrucciones, datos, etc... Por ejemplo, una lista de instrucciones y datos constituye un **programa**. Los programas que realizan determinadas tareas se llaman **rutinas**.

La unidad aritmética debe ser capaz de efectuar las operaciones que se le ordenen desde la unidad de control. También puede tener una memoria propia, donde se escriben los datos y resultados.

El dispositivo de salida es un sistema de presentación de datos, como una pantalla, una impresora, etc... También puede ser una unidad periférica que enlace con procesos exteriores, incluso que incluya una adaptación al **tiempo real** exterior.

13.17.- Lenguajes. Informática.

En los ejemplos anteriores se comprende que las instrucciones u órdenes de actuación vendrán codificadas en **lenguaje de máquina,** es decir, de forma que el ordenador las pueda interpretar y serán una sucesión de unos y ceros dentro del programa. En algunos ordenadores se dan las instrucciones y los datos por vías distintas y en otros se dan por la misma vía y el ordenador distingue las instrucciones y los datos por su codificación y posición dentro del programa. Un **programa en lenguaje de máquina** es un conjunto de bits agrupados en bytes (por ejemplo, en grupos de 16 o 32 bits) que van significando instrucciones y datos exactamente colocados en el sitio que les corresponde. Antiguamente (en técnicas de teletipos) los bits se agrupaban en "caracteres" de unos 8 bits y éstos en "palabras" de 6 o más caracteres. Actualmente, la definición de carácter ha desaparecido como tal grupo de bits, sustituida por el "byte", y la definición de palabra es algo confusa y dependiente del fabricante, pero dominan los conceptos de bit y byte.

En última instancia siempre hay que dar al ordenador el programa en lenguaje de máquina. Pero la mayoría de los ordenadores poseen un dispositivo lógico para "traducir" o "codificar" en lenguaje de máquina instrucciones y datos escritos de forma más comprensible para el operador. Según lo elaborado que es esta codificación, se dice que esta forma de escribir o lenguaje tiene más o menos **nivel.**

Si las instrucciones se escriben según un catálogo de instrucciones sencillo, como, por ejemplo, LDA, que significa cargar en el registro A lo que venga detrás, el lenguaje se llama **assembly o assembler.** El codificador que convierte estas instrucciones sencillas en lenguaje de máquina se llama también assembler. Cuando el ordenador hace automáticamente estas codificaciones se dice que tiene un **editor.**

Existen lenguajes más elaborados e **independientes del ordenador** donde las operaciones están designadas en forma universal, como, por ejemplo, "add" para sumar. Estos lenguajes se "pasan" a lenguaje de máquina mediante un editor automático (que sí depende del ordenador) que se llama **compiler o compilador.** Ejemplos de estos lenguajes de "alto nivel" son el **FORTRAN** (formule translation...), adecuado para las matemáticas, COBOL (common business oriented lenguage), **BASIC** (beginners algebraic symbol interpreter compiler), etc... Algunos de estos lenguajes permiten interrumpir las rutinas y "dialogar" con el

ordenador y entonces se llama **interpretativo.** Son los más usados en calculadoras directas o de bolsillo por su fácil utilización. También surgen lenguajes más elaborados de más nivel, como **PASCAL**, **UNIX**, **ADA**, con usos específicos y los llamados **sistemas operativos,** como **MSDOS** o **WINDOWS**, que permiten realizar funciones más "inteligentes", como **procesar textos, controlar comunicaciones,** etc...

La ciencia de manejo de todos estos recursos, llamados también de **software** ha nacido con el nombre de **Informática,** y se aleja de la Electrónica, que es el propósito de este libro, por lo que, para profundizar en ella, se invita al lector a acudir a la literatura de Informática. Frente al "software" ha surgido la tecnología de proyecto de base electrónica (circuitos, etc...) que se llama de **hardware** y que ha sido el objeto de este capítulo.

13.18.- Microcomputadores y microprocesadores.

En los últimos años se ha extendido el uso de computadores de un alto grado de integración. Se suele llamar **microcomputadores** o **microprocesadores** a los ordenadores construidos sobre la base de una pastilla llamada **CPU**, (Central Processor Unit) que incluye la unidad central de control, unidad aritmética y una parte de la memoria.

La figura 13-23 representa el esquema básico de un microprocesador elemental. Un reloj inicia y mantiene el ritmo de un contador de programa que va leyendo las posiciones sucesivas de una memoria usando la salida del contador como "direccionamiento" para la memoria.

El contenido de las posiciones sucesivas de la memoria se va interpretando en un "decodificador de instrucciones", que regula la secuencia de operaciones, o bien entra como datos en la unidad aritmética. Los resultados de las operaciones se "extraen" de nuevo hacia la memoria (a posiciones distintas del programa) y se sigue el programa. Los periféricos se "direccionan" como posiciones de memoria y dan acceso al mundo exterior.

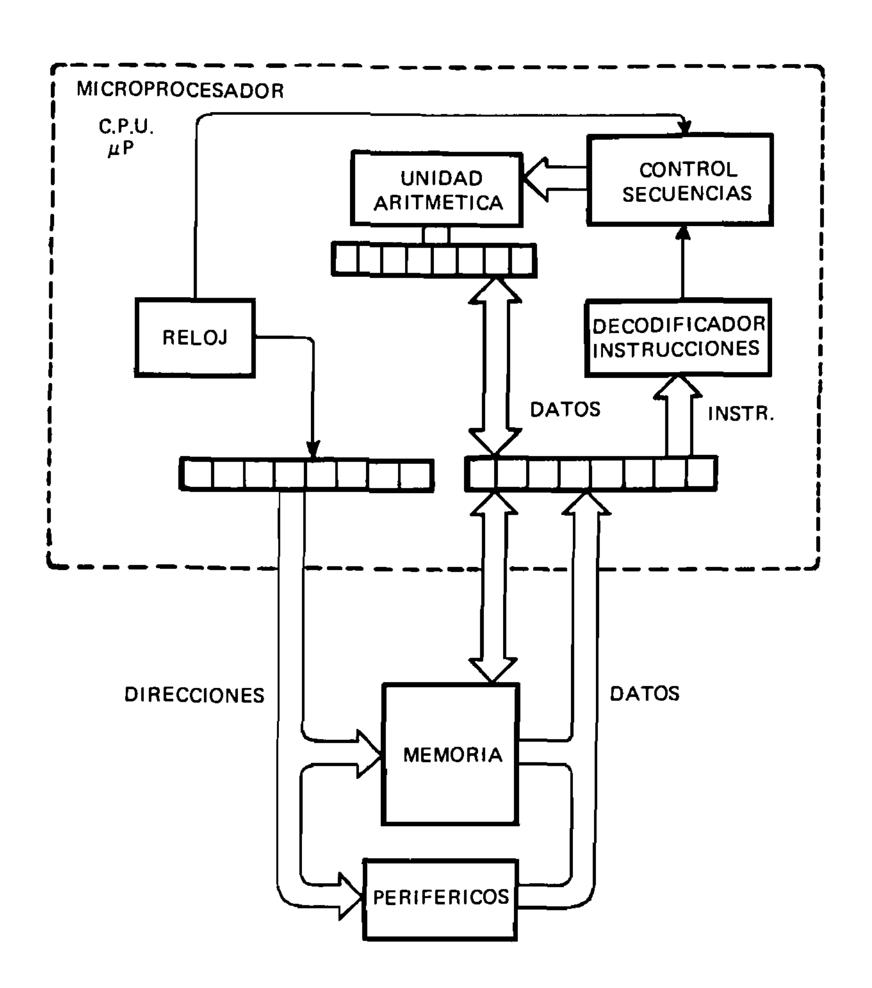

MICROPROCESADOR
C.P.U.
μP
UNIDAD ARITMETICA
CONTROL SECUENCIAS
RELOJ
DECODIFICADOR INSTRUCCIONES
DATOS
INSTR.
DIRECCIONES
DATOS
MEMORIA
PERIFERICOS

Fig. 13-23.

CAPITULO 14

LINEAS DE TRANSMISION Y GUIAS DE ONDA

14.1.- Introducción.

En los capítulos anteriores hemos estudiado el comportamiento de circuitos con **constantes concentradas**, es decir, construidos con componentes conectados entre sí. En ellos hemos ignorado, considerándolo despreciable, el tiempo de tránsito por las conexiones. Sin embargo, es corriente el uso de **líneas de transmisión**, que, en un principio se usaron para transmitir señales de un sitio a otro y, posteriormente se usaron también por sus propiedades especiales en cuanto a resonancias, desfases, etc..., como componentes especiales de **constantes distribuidas**. Estos dispositivos solo necesitan un estudio o consideración especial, como el de este capítulo, cuando su longitud es grande comparada con la longitud de onda de las señales que se transmiten. Si no es así, pueden tratarse como simples conexiones, ignorando el tiempo de transmisión. Estos efectos se manifiestan solamente a frecuencias suficientemente altas (longitudes de onda cortas), por lo que a veces se les llama "líneas de VHF". Por encima de esas frecuencias se usan las **guías de onda**, dispositivos de ondas electromagnéticas que también estudiaremos en este capítulo.

14.2.- Líneas de transmisión.

La figura 14-1 es el esquema de una línea de transmisión. La línea, en este ejemplo, se compone de dos conductores paralelos que unen el **generador** con la **carga**.

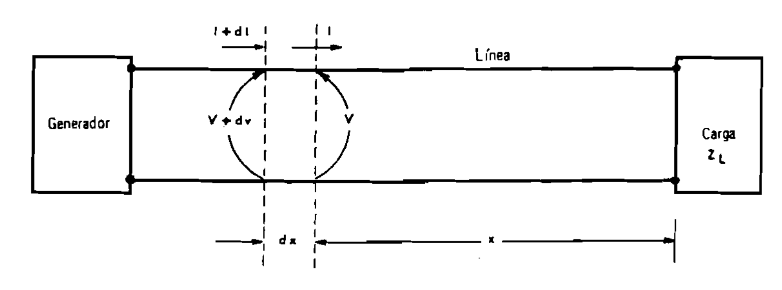


Fig. 14-1.

A partir de cierta frecuencia, las líneas formadas por dos conductores paralelos radian gran cantidad de energía (cuando la distancia entre los conductores es comparable con la longitud de onda). Las líneas así construidas dejan de ser utilizables, por sus elevadas pérdidas. A altas frecuencias, es conveniente utilizar líneas **coaxiales**, formadas por un conductor central rodeado por una malla metálica de forma tubular que hace de blindaje a potencial casi constante y forma el "retorno de tierra" de muy baja impedancia. Los dos conductores (central y malla) están separados por un dieléctrico apropiado. Las constantes distribuidas a lo largo de la línea son:

Resistencia óhmica en serie por unidad de longitud **R**.

Autoinducción en serie por unidad de longitud **L**.

Capacidad paralela distribuida por unidad de longitud **C**.

Conductancia en paralelo por unidad de longitud **G**.

Las constantes distribuidas pueden agruparse en el elemento dx de la línea, de la forma que se representa en la figura 14-2. Si la línea es paralela, los valores en serie pueden suponerse divididos en dos mitades, una parte como está en la figura y la vuelta en el conductor de retorno, pero las ecuaciones son las mismas con alguna ligera aproximación, por lo que sólo trataremos las líneas coaxiales.

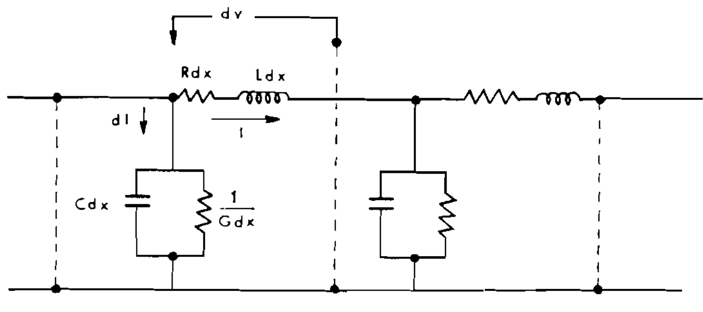


Fig. 14-2.

Las ecuaciones de la red de la figura son (suponiendo ondas sinusoidales y utilizando el operador j ω) :

$$dV=(Rdx+j\omega Ldx)\,I=(R+j\omega L)\,Idx$$
$$dI=(Gdx+j\omega Cdx)\,V=(G+j\omega C)\,Vdx$$

Si llamamos Z a la impedancia compleja serie por unidad de longitud (paréntesis de la primera ecuación) e Y a la admitancia compleja paralelo por unidad de longitud (paréntesis de la segunda ecuación), las ecuaciones se simplifican y quedan:

$$\frac{dV}{dx}=ZI \quad ; \quad \frac{dI}{dx}=YV$$

Derivando la primera y sustituyendo en la segunda se obtiene la "ecuación de propagación" de segundo orden :

$$\frac{d^2V}{dx^2}=ZYV$$

La solución de esta ecuación es:

$$V=V_1e^{\sqrt{ZY}x}+V_2e^{-\sqrt{ZY}x}$$

donde V_1 y V_2 son constantes que dependen de las condiciones en los límites, que aquí serán el **generador** y la **carga**.

Derivando y dividiendo por Z (de la primera ecuación de la línea) se obtiene la intensidad :

$$I=\frac{V_1}{Z}\sqrt{ZY}e^{\sqrt{ZY}x}-\frac{V_2}{Z}\sqrt{ZY}e^{-\sqrt{ZY}x}=$$
$$I=\frac{V_1}{\sqrt{\frac{Z}{Y}}}e^{\sqrt{ZY}x}-\frac{V_2}{\sqrt{\frac{Z}{Y}}}e^{-\sqrt{ZY}x}$$

La constante $\sqrt{ZY}$ se llama **constante de propagación**, que es un complejo **α + βj**. La parte real se llama **atenuación** y la parte imaginaria **constante de fase**. El valor de estas constantes se calcula a partir de la igualdad:

$$\alpha+\beta j=\sqrt{ZY}=\sqrt{(R+j\omega L)\,(G+j\omega C)}$$

Teniendo en cuenta que $\omega C>>G$ y $\omega L>>R$, se calcula

directamente el valor:

$$\beta=\omega\sqrt{LC}$$

Para hallar α se eleva al cuadrado:

$$RG-\omega^2LC+j\omega CR+j\omega LG=\alpha^2-\beta^2+2\alpha\beta j$$

Igualando las partes imaginarias y sustituyendo el valor aproximado de β :

$$\alpha=\frac{\omega CR+\omega LG}{2\beta}=\frac{1}{2}(R\sqrt{\frac{C}{L}}+G\sqrt{\frac{L}{C}})$$

La constante

$$Z_0=\sqrt{\frac{Z}{Y}}=\sqrt{\frac{R+j\omega L}{G+j\omega C}}\approx\sqrt{\frac{L}{C}}$$

se llama **impedancia característica** de la línea.

14.3.- Interpretación de la solución.

La solución en tensión e intensidad se compone de **dos ondas**:

Una **onda directa** de tensión

$$V_1e^{\sqrt{ZY}x}=V_1e^{\alpha x}e^{\beta jx}$$

cuyo valor aumenta en fase β radianes por unidad de longitud **desde la carga**, es decir, se retrasa hacia la carga. Recuérdese que hemos tomado el origen de x en la carga, creciendo x hacia el generador. El retrasarse hacia la carga podemos interpretarlo como que se **propaga desde el generador hacia la carga**. Además, se atenúa hacia la carga a razón de e^{α} por unidad de longitud, o, puede decirse, aumenta e^{α} hacia el generador.

Recordemos que en todos los cálculos anteriores hemos usado la transformación compleja j ω. En particular, la onda directa obtenida es la transformada compleja de una función del tiempo de la misma amplitud y fase, es decir, que la onda directa en el plano del tiempo será:

$$V_d = V_1 e^{\alpha x} \cos(\omega t + \beta x)$$

De esta ecuación se deduce el valor de la **longitud de onda en la línea** λ_l, incremento de x para que se repita el valor de fase, que es:

$$\lambda_l = \frac{2\pi}{\beta} = \frac{2\pi}{\omega\sqrt{LC}} = \frac{1}{f\sqrt{LC}}$$

La velocidad de avance de la onda o **velocidad de fase** v_l se deduce de la relación entre x y t para fase constante. O sea que la velocidad de fase en la línea será (tomando x con signo - para expresar que es velocidad hacia el generador, x decreciente:

$$\omega t + \beta x = cte., \quad \omega dt + \beta dx = 0$$
$$V_l = -\frac{dx}{dt} = \frac{\omega}{\beta} = \frac{2\pi f}{\beta} = f\lambda = \frac{1}{\sqrt{LC}}$$

El cálculo de L y C en función de μ y ϵ da un valor próximo al de avance en el vacío, es decir, la velocidad de la luz. La longitud de onda en la línea es ligeramente inferior a la del vacío.

A esta onda de tensión directa le corresponde una onda de **intensidad directa**, que vale:

$$\frac{V_1 e^{\sqrt{ZY}x}}{Z_0}$$

es decir, la tensión de la onda directa dividida por la impedancia característica.

Además existe una **onda reflejada** de amplitud:

$$V_2 e^{-\sqrt{ZY}x} = V_2 e^{-\alpha x} e^{-\beta j x}$$

creciente hacia la carga (decreciente hacia el generador) y que se retrasa al crecer x (hacia el generador), lo que justifica la interpretación como **"reflejada"** desde la carga hacia el generador.

A esta onda de tensión corresponde una onda de **intensidad reflejada:**

$$\frac{V_2 e^{-\sqrt{ZY}x}}{Z_0}$$

cuyo signo indica que está "generada" a la derecha (reflejada desde la carga) y que es igual a la tensión reflejada dividida por la impedancia característica.

Ambas ondas, directa y reflejada **ven** pues la misma impedancia, la impedancia característica, de valor:

$$Z_0=\sqrt{\frac{L}{C}}$$

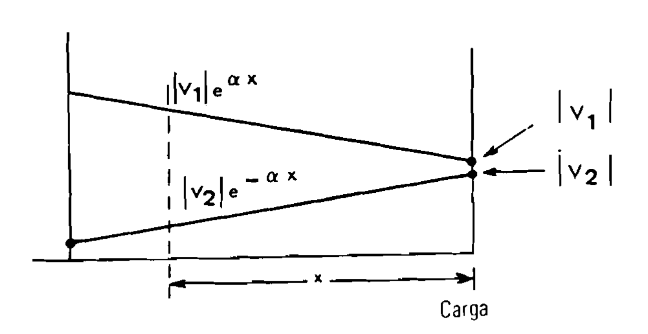


Fig. 14-3.

La figura 14-3 representa las amplitudes de las ondas directa y reflejada a lo largo de la línea. Recordemos que V_1 y V_2 son complejos con módulo y argumento, que representan ondas sinusoidales del tiempo.

14.4.- Ecuación de carga.

La solución de las ecuaciones, que hemos obtenido en el párrafo anterior, deja **dos** constantes de integración indeterminadas, V_1 y V_2, que se hallarán con las **condiciones en los límites**. Normalmente, la ecuación de carga fija una relación entre ellas y su amplitud viene determinada finalmente por el generador.

Como el problema es lineal, normalmente la ecuación de carga permite calcular la **impedancia reflejada** para cada valor de x, y, con la tensión del generador y la impedancia reflejada a la altura del generador, se calcula la intensidad.

Para x=0 los valores de V e I son:

$$V=V_1+V_2$$
$$I=\frac{V_1}{Z_0}-\frac{V_2}{Z_0}$$

La ecuación de carga con impedancia Z_L conectada es:

$$Z_L=\frac{V}{I}=\frac{V_1+V_2}{V_1-V_2}Z_0$$
$$o\ sea$$
$$\frac{V_1+V_2}{V_1-V_2}=\frac{Z_L}{Z_0}$$

Al valor Z_L / Z_0 se le llama **carga reducida**. El valor de la onda reflejada dividida por la directa

$$\rho=\frac{V_2}{V_1}$$

se le llama **coeficiente de reflexión.** Su valor se obtiene de la ecuación anterior, dividiendo el numerador y denominador por V_2 y despejando el valor del coeficiente de reflexión:

$$\frac{1+\rho}{1-\rho}=\frac{Z_L}{Z_0}$$
$$o\ sea$$
$$\rho=\frac{V_2}{V_1}=\frac{Z_L-Z_0}{Z_L+Z_0}$$

Cuando la línea termina en su impedancia característica, o sea, $\mathbf{Z_L = Z_0}$, el coeficiente de reflexión es $\rho = 0$, **no hay reflexiones**, se dice que la línea está "bien terminada".

Si la impedancia terminal es muy distinta de la característica, el coeficiente de reflexión no es cero y hay reflexiones. Por ejemplo, si Z_L =0, línea **cortocircuitada**, ρ=-1 y se refleja V_2 = - V_1, es decir se refleja **totalmente** la onda incidente **cambiada de signo**.

Si $Z_L = \infty, \rho = 1$ y se refleja $V_2 = V_1$, es decir, la onda incidente se refleja **totalmente, sin cambio de signo.**

14.5.- Ondas estacionarias.

En cualquier punto x de la línea, la tensión y la intensidad valen (sustituyendo $V_2 = \rho V_1$):

$$V=V_1\,[e^{\sqrt{ZY}x}+\rho\, e^{-\sqrt{ZY}x}]$$
$$I=\frac{V_1}{Z_0}\,[e^{\sqrt{ZY}x}-\rho\, e^{-\sqrt{ZY}x}]$$

donde Z,Y (y, por consiguiente Z_0) son parámetros **de la línea**, ρ depende de la impedancia de carga y V_1 se calcula con la tensión del generador para el valor de x igual a la longitud de la línea. La fase de la onda directa aumenta hacia el generador y la de la reflejada hacia la carga, a razón de β radianes por metro.

Si llamamos ρ_0 y γ al módulo y argumento de ρ, los valores de tensión e intensidad (cada una de ellas suma de directa y reflejada) a lo largo de la línea valen:

$$V=V_1\,[e^{\alpha x}e^{\beta xj}+\rho_0 e^{-\alpha x}e^{(\gamma-\beta x)j}]$$
$$I=\frac{V_1}{Z_0}\,[e^{\alpha x}e^{\beta xj}-\rho_0 e^{-\alpha x}e^{(\gamma-\beta x)j}]$$

Las ondas de tensión directa y reflejada se pueden representar por vectores cuyas amplitudes son $V_1\, e^{\alpha x}$ y $V_1\, \rho_0\, e^{-\alpha x}$ y cuyas fases son βx y γ - βx, representados en la figura 14-4.

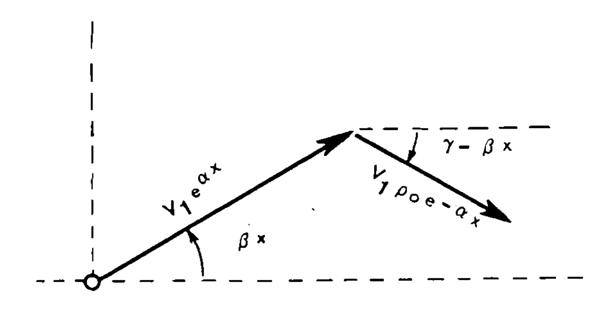


Fig. 14-4.

El primero de estos vectores, onda directa de tensión, aumenta con x y el segundo, onda reflejada de tensión, disminuye con x. La fase **relativa** de estos dos vectores es $2\beta x - \gamma$. Cada vez que esta fase pasa por $2\pi n$, es decir, a distancias en que βx aumenta π, estos dos vectores están **en fase**, y cada vez que pase por $(2n + 1)\pi$ (en los puntos intermedios) los dos vectores están en **contrafase**, se restan. Al valor :

$$\lambda_x = \frac{2\pi}{\beta} = \frac{2\pi}{\omega\sqrt{LC}} = \frac{\lambda}{c\sqrt{LC}}$$

se le llama **longitud de onda de la línea**. En la práctica suele valer 0,6...0,9 λ.

Así pues, si medimos la distancia en "ángulos de línea" en radianes, x/β, cada distancia π, es decir, 180 º **se suman la onda directa y reflejada dando un máximo (vientre)**. En los puntos intermedios, a 90º de los anteriores, se **resta** la reflejada de la directa, dando un **mínimo (nodo)**. Esta interferencia entre onda directa y reflejada se conoce como **onda estacionaria**, llamada así porque los vientres y nodos están **en puntos fijos de la línea.**

El mismo esquema podría hacerse con la onda de intensidad. Debido al signo menos de la onda reflejada, **los vientres de la onda de intensidad corresponden a los nodos de la de tensión y viceversa.**

14.6.- Valores reflejados.

El valor del **coeficiente de reflexión** ρ_x, relación entre la tensión directa y la reflejada (los complejos que las representan) en un punto de la línea a distancia x de la carga es:

$$\rho_x = \frac{V_1 \rho e^{-\sqrt{ZY}x}}{V_1 e^{\sqrt{ZY}x}} = \rho e^{-2\sqrt{ZY}x}$$

El valor de la **impedancia reflejada**, V/I en cualquier punto de la línea vale, en función de x:

$$Z_x = \frac{V}{I} = \frac{V_1 (e^{\sqrt{ZY}x} + \rho e^{-\sqrt{ZY}x})}{\frac{V_1}{Z_0}(e^{\sqrt{ZY}x} - \rho e^{-\sqrt{ZY}x})} =$$

$$= Z_0 \frac{(Z_L+Z_0)\, e^{\sqrt{ZY}x} + (Z_L-Z_0)\, e^{-\sqrt{ZY}x}}{(Z_L+Z_0)\, e^{\sqrt{ZY}x} - (Z_L-Z_0)\, e^{-\sqrt{ZY}x}} =$$

$$= Z_0 \frac{Z_L \cosh\sqrt{ZY}x + Z_0 senh\sqrt{ZY}x}{Z_L senh\sqrt{ZY}x + Z_0 \cosh\sqrt{ZY}x}$$

$$Z_0 \frac{Z_L + Z_0 tgh\sqrt{ZY}x}{Z_0 + Z_L tgh\sqrt{ZY}x}$$

14.7.- Líneas sin pérdidas.

En muchos casos los valores de R y G son despreciables frente a ωL y ωC y las fórmulas se simplifican mucho. En efecto, en ese caso, se verifica:

$$\sqrt{ZY}=j\omega\sqrt{LC} \text{ y } \alpha=0$$

La impedancia característica es una resistencia pura:

$$Z_0=\sqrt{\frac{L}{C}}$$

El coeficiente de reflexión vale, como antes:

$$\rho=\frac{Z_L-Z_0}{Z_L+Z_0}$$

La simplificación más importante es que las ondas directa y reflejada **no se atenúan**. Los valores de V e I son :

$$V=V_1\,(e^{j\beta x}+\rho e^{-j\beta x})$$
$$I=\frac{V_1}{Z_0}\,(e^{j\beta x}-\rho e^{-j\beta x})$$

La tensión pasa por **máximos** y **mínimos** de valor:

$$V_{max}=|V_1|\,(1+|\rho|)$$
$$V_{min}=|V_1|\,(1-|\rho|)$$

El coeficiente

$$\rho=\frac{Z_L-Z_0}{Z_L+Z_0}$$

tiene módulo siempre inferior a 1 como es fácil demostrar por ser Z_0 resistivo y la parte real de Z_0 positiva.

Los máximos y mínimos de la intensidad valen también :

$$I_{max}=\frac{|V_1|}{Z_0}(1+\rho)$$
$$I_{min}=\frac{|V_1|}{Z_0}(1-\rho)$$

La relación entre máximos y mínimos, de tensión o intensidad, se llama **relación de ondas estacionarias** y vale:

$$S=\frac{|V_1|(1+|\rho|)}{|V_1|(1-|\rho|)}$$

La **impedancia reflejada** en función de x vale:

$$Z_x=Z_0\frac{Z_L+jZ_0 tg\beta x}{Z_0+jZ_L tg\beta x}$$

El **coeficiente de reflexión en el punto x**, definido como relación entre la onda reflejada y la directa en ese punto vale:

$$\rho_x=\rho e^{-2j\beta x}$$

14.8.- Línea abierta y en cortocircuito.

Una línea se llama **abierta** cuando la impedancia de carga es infinito, $Z_L \to \infty$, en cuyo caso el coeficiente de reflexión es 1 y los valores de tensión e intensidad a lo largo de la línea son:

$$V=V_1(e^{j\beta x}+e^{-j\beta x})=2V_1\cos\beta x$$
$$I=\frac{V_1}{Z_0}(e^{j\beta x}-e^{-j\beta x})=\frac{2V_1}{Z_0}jsen\beta x$$

La amplitud de las ondas estacionarias de tensión pasa por un máximo de $2V_1$ cada vez que x aumenta π/β, es decir, cada **media longitud de onda de la línea**, en los puntos en que βx es múltiplo de π. En esos puntos **V es real**, lo que quiere decir que la fase de la onda estacionaria es siempre 0 o 180º, alternativamente (según que el coseno valga +1 o -1).

En esos puntos (máximo de la tensión), la **intensidad es 0.** La intensidad pasa por máximos (con el seno de βx) en los puntos

de tensión 0. La intensidad está siempre en cuadratura con la tensión (factor j), pero tiene la misma fase dentro de cada "lóbulo". La figura 14-5 representa las distribuciones de tensión e intensidad en una línea "abierta" en su extremo (sin carga).

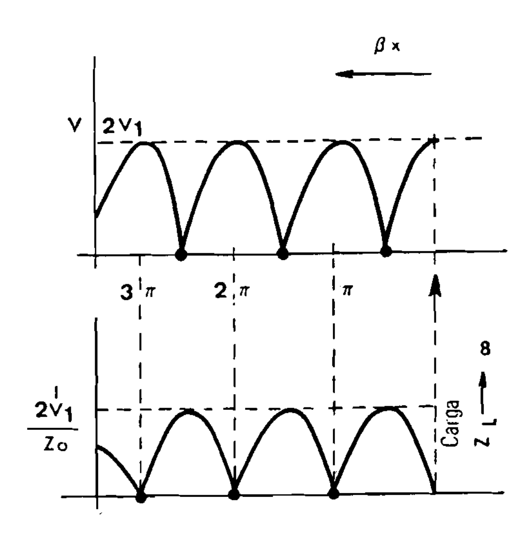


Fig. 14-5.

La **impedancia reflejada** V/I vale en función del punto x:

$$Z_x=\frac{V}{I}=-Z_0 j\ ctg\ \beta x$$

y se hace ∞ en los vientres y 0 junto con la tensión. Entre los valores 0 y ∞/2 de x la impedancia toma todos los valores de reactancia (imaginaria pura), de ∞ a 0.

Si la línea está **cortocircuitada,** Z_L = 0 y se tiene:

$$\rho=\frac{Z_L-Z_0}{Z_L+Z_0}=-1$$
$$V=V_1\,(e^{j\beta x}-e^{-j\beta x})=2V_1 j sen\beta x$$
$$I=\frac{V_1}{Z_0}\,(e^{j\beta x}+e^{-j\beta x})=\frac{2V_1}{Z_0}\cos\beta x$$
$$Z_x=+Z_0 j tg\beta x$$

La amplitud de las ondas estacionarias de tensión es 0 en los múltiplos de media longitud de onda y máxima e igual a $2V_1$ en los puntos intermedios. La intensidad está en cuadratura, tanto respecto a x como en cada punto. La impedancia es cero para x =0 y distancias múltiplos de media longitud de onda, y se hace ∞ en

los puntos intermedios. Como la impedancia $Z_L = 0$ se refleja como $Z_L = \infty$ a distancia de un cuarto de longitud de onda, la línea cortocircuitada puede considerarse como una línea abierta a partir de este punto. La figura 14-6 representa las ondas estacionarias de tensión e intensidad en una línea cortocircuitada.

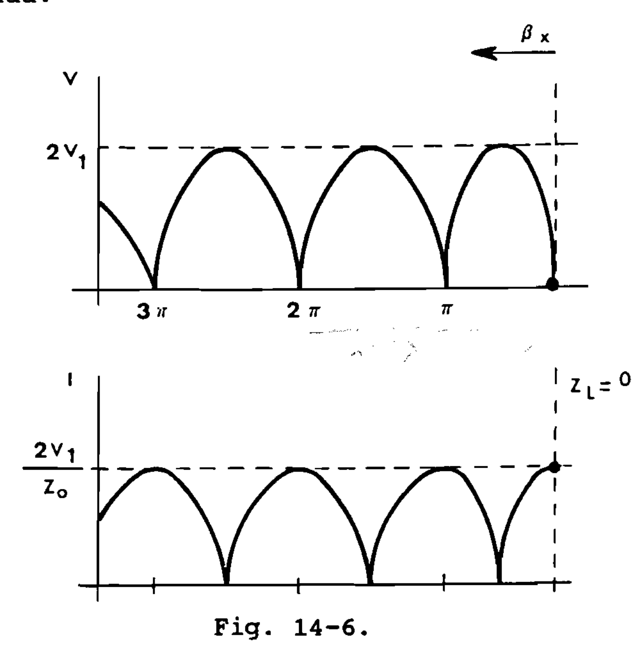


Fig. 14-6.

14.9.- Transformador de cuarto de onda.

La propiedad de variar la impedancia en las líneas se utiliza para **adaptar** impedancias. En efecto si cortamos una línea a la longitud **cuarto de longitud de onda**, es decir a un valor de $x=l$ tal que, $\beta l = \pi / 2$, y la terminamos con una impedancia Z_L, se tiene una impedancia reflejada que vale:

$$Z=Z_0\frac{Z_L+jZ_0 tg\beta l}{Z_0+JZ_L tg\beta l}=\frac{Z_0^2}{Z_L}$$

Esta fórmula permite reflejar una impedancia cualquiera a partir de otra Z_L, ambas resistivas (reales puras). Por ejemplo, si se quiere obtener una resistencia Z a partir de otra Z_L se puede hacer con una línea de impedancia característica:

$$Z_0=\sqrt{Z_L Z}$$

En el caso de impedancias complejas, Z_L y Z deben dar un producto real, para que Z_0 resulte real.

14.10.- Líneas terminadas en su impedancia característica.

Si $Z_L = Z_0$, la línea está "bien terminada", ρ=0 y no hay ondas estacionarias. Los valores a lo largo de la línea son:

$$V=V_1 e^{j\beta x}$$
$$I=\frac{V_1}{Z_0} e^{j\beta x}$$
$$Z_x=\frac{V}{I}=Z_0$$

La impedancia reflejada en cualquier punto es siempre 0 y las amplitudes de V e I son constantes.

La propiedad de que en una línea terminada en su impedancia característica **no hay reflexiones** es muy importante ya que lo que se pretende con la línea es que la señal viaje del generador a la carga sin deterioro, tal y como ocurre en estas líneas. Aunque las ecuaciones que hemos aplicado eran sólo válidas para señales sinusoidales, es fácil generalizar estas últimas conclusiones a otras señales, como digitales, o con un espectro variado, es decir, que **terminando la línea en su impedancia característica se evitan las reflexiones** (a cualquier frecuencia). Además, la eliminación de ondas estacionarias evita otros fenómenos como sobretensiones etc... También la longitud de la línea deja de ser crítica en cuanto posibles adaptaciones de impedancia.

Normalmente, las impedancias características son de unos 50 a 100 ohmios para las líneas coaxiales y de 150, 300 y 600 ohmios para las paralelas.

Muy importantes son las reflexiones producidas por cambios o irregularidades en la línea. Por eso es imprescindible, para evitar discontinuidades de la impedancia, utilizar conectores especiales, llamados "de UHF", que mantienen (aproximadamente) la impedancia característica a través de las conexiones.

14.11.- Diagrama de Smith.

El diagrama de Smith es un ábaco ingenioso que permite efectuar rápidamente todos los cálculos de impedancias de las **líneas sin pérdidas**.

Su fundamento es el siguiente: Teniendo en cuenta que el módulo del coeficiente de reflexión es menor que 1, en el "plano" ρ estarán representados todos los valores del coeficiente de reflexión dentro de un círculo de radio unidad. El **diagrama de Smith** es una representación de los lugares geométricos de parte real y parte imaginaria de la impedancia reducida Z_x / Z_0 constante en el plano ρ. Teniendo en cuenta la relación:

$$\rho=\frac{Z-Z_0}{Z+Z_0}=\frac{Z/Z_0-1}{Z/Z_0+1}=1-\frac{2}{Z/Z_0+1}$$

y tomando esta última relación como una transformación compleja del plano Z/Z_0 al plano ρ, la transformación compleja se descompone en (llamando $z = Z/Z_0$):

Una traslación +1 de una unidad hacia la derecha, obteniendo z+1.

Una inversión, obteniendo 1/(z+1).

Una homotecia, multiplicando por 2, obteniendo 2/(z+1).

Una simetría sobre el eje vertical, obteniendo - 2/(z+1).

Una traslación final +1 de una unidad hacia la derecha.

Los lugares geométricos de los valores de z real, que eran líneas verticales en el plano Z, se convierten en círculos tangentes al eje vertical en la inversión que luego pasan a círculos tangentes a la vertical que pasa por el punto 1,0.

Los lugares geométricos de los valores de z imaginarios puros forman una familia de círculos ortogonal a la anterior.

En definitiva, el diagrama queda como se ve en la figura 14-7, formando el **ábaco de Smith**.

En la parte exterior del círculo de radio 1 se señalan los valores del argumento del coeficiente de reflexión en grados, a partir de un punto en que éste sea real . Recuérdese que el valor del coeficiente de reflexión es:

$$\rho_x=\rho e^{-2j\beta x}$$

con lo que los valores del argumento son **el doble de la "longitud eléctrica" de la línea, en grados**. La longitud eléctrica **en longitudes de onda de la línea** está también indicada, correspondiendo 180º de coeficiente de reflexión a 90º de longitud eléctrica, es decir 0,25 de la longitud de onda.

El círculo exterior corresponde a módulo 1 del coeficiente de reflexión y los círculos concéntricos **no dibujados** corresponden a otros valores. **Todos los valores de la línea están en un círculo concéntrico**, ya que el módulo del coeficiente de reflexión no varía a lo largo de la línea.

En las dos familias de círculos ortogonales dibujadas están marcados los valores de la parte real **resistiva** y la imaginaria **reactiva** de z_x.

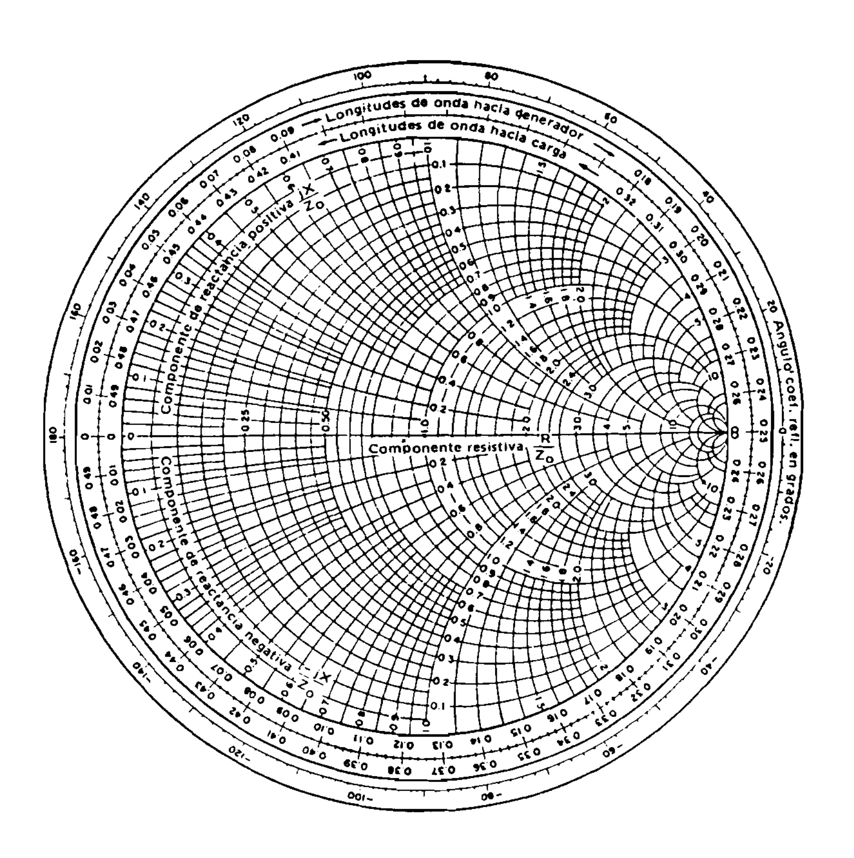


Fig. 14-7.

Veamos ahora como se maneja el diagrama:

Dado un valor $z = Z_L/Z_0$ de **carga** se busca el punto correspondiente por intersección de los círculos de parte real y parte imaginaria de z. A partir de este punto, la **línea** está

representada por un **círculo concéntrico** con el exterior, que hay que dibujar con un compás. El coeficiente de reflexión ρ tiene por módulo el **radio** de ese círculo y por argumento el que se lee prolongando el radio hasta el exterior. A partir de ese punto se gira por ese círculo un arco de círculo en dirección **hacia el generador** (señalada en el diagrama) de magnitud el doble de la longitud eléctrica de la línea hasta el punto $\mathbf{x}$, es decir, $2\beta x$ en ángulo (señalado en la periferia) o en fracciones de longitud de onda de la línea, también señaladas. Medio círculo, 180º o π radianes, corresponde a un cuarto de longitud de onda, porque el ángulo del coeficiente de reflexión es el doble de la longitud eléctrica. Con ello se llega a un nuevo punto cuyas partes real e imaginaria pueden leerse en las familias ortogonales del diagrama.

El problema descrito es el que permite calcular la impedancia reflejada en cualquier punto x de la línea, pero, naturalmente, el problema puede invertirse, calculando, por ejemplo, el ángulo que hay que girar (es decir, la longitud que debe tener la línea) para obtener una impedancia resistiva o de parte real determinada, etc...

Las líneas abiertas o cortocircuitadas están representadas en el círculo exterior, porque la impedancia de carga es 0 o ∞.

Debido a la forma de representación, los puntos simétricos respecto al centro representan **complejos inversos**. Esta propiedad permite utilizar el diagrama como **diagrama de admitancias**, por el mismo procedimiento: las líneas están representadas por círculos concéntricos, no dibujados, cuyos argumentos son distancias hacia el generador o la carga. Los puntos son **admitancias**, $\mathbf{Z_0/Z_L}$ en cada punto. De admitancias a impedancias o viceversa se pasa por simetría respecto al origen.

14.12.- Adaptación por medio de un fragmento de línea.

Un método muy usado de adaptar una impedancia terminal distinta de la característica es colocar en el punto adecuado de la línea un fragmento de línea en paralelo que modifica la impedancia reflejada.

Por ejemplo, en la figura 14-8, supongamos que la impedancia de carga Z_L es diferente de la característica Z_0 . Se trata de elegir las longitudes x_1 y x_2 de forma que la impedancia Z_1 (ver figura 14-9) en paralelo con la Z_2 dé la impedancia característica Z_0.

Para ello se utiliza el diagrama de Smith de la siguiente forma:

Se busca Z_L/Z_0 en el diagrama y se **pasa a admitancia** tomando el punto simétrico respecto al centro. A partir de ese punto se describe un arco de círculo concéntrico (no dibujado) en dirección hacia el generador (sentido de las agujas del reloj) hasta encontrar el círculo de **parte real uno**. El arco descrito es la longitud de la línea que puede leerse en **longitudes de**

onda. Si se lee en grados hay que recordar que hay que dividir por 2, porque los arcos del diagrama son $2\beta x$. En ese punto la admitancia es Z_0/Z_l y la parte real es la correcta, 1. Ahora hay que cortar la línea abierta en paralelo para que refleje una componente imaginaria igual y contraria a la parte imaginaria de ese punto. Por consiguiente se lee la parte imaginaria de la admitancia Z_0/Z_l y se busca el valor igual y contrario (simétrico respecto al eje horizontal) y se busca ese valor (siguiendo el círculo de reactancia constante)en el círculo exterior, que es el que representa la línea abierta. A continuación se recorre este círculo exterior en dirección hacia la carga hasta llegar a 0 (admitancia 0, impedancia infinito, que corresponde al final de la línea abierta). El arco recorrido será el que corresponde a x_2 en longitudes de onda.

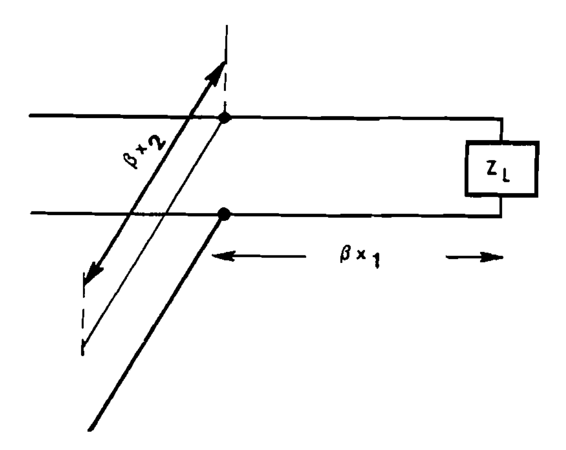


Fig. 14-8.

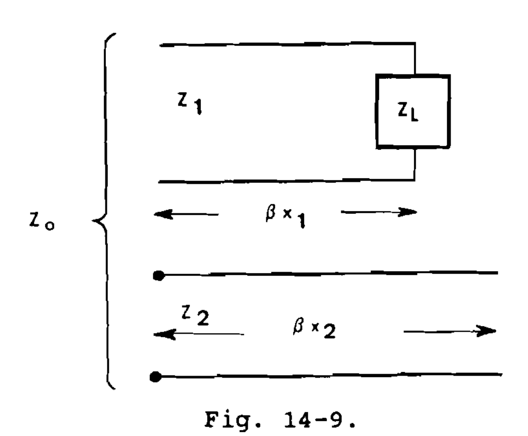


Fig. 14-9.

14.13.- Ondas electromagnéticas. Ecuaciones de Maxwell.

Las ondas electromagnéticas son un vehículo ideal para enviar información a través del espacio. Las ondas electromagnéticas son un mecanismo de interacción entre el campo

eléctrico y el magnético que se **propaga** a través de medios de naturaleza diversa y, en particular, a través del vacío o del aire.

La propagación de ondas electromagnéticas obedece a las ecuaciones de Maxwell, que son:

$$\nabla\vec{E}=0$$
$$\nabla\vec{H}=0$$
$$\nabla\times\vec{H}=\epsilon\frac{\partial\vec{E}}{\partial t}$$
$$\nabla\times\vec{E}=-\mu\frac{\partial\vec{H}}{\partial t}$$

donde los vectores E y H representan los campos eléctrico y magnético, ϵ y μ son la constante dieléctrica y la permeabilidad del medio.

Las dos primeras ecuaciones expresan que la divergencia del campo eléctrico y del magnético son cero, es decir, que no hay cargas eléctricas ni magnéticas en el medio.

La tercera ecuación expresa que la variación del campo eléctrico produce un campo magnético alrededor del que es el rotacional y es una generalización de la ley de Ampere.

La cuarta ecuación expresa que la variación del campo magnético genera un campo eléctrico "inducido" del que es el rotacional y es una generalización de la ley de Faraday.

Las ecuaciones de Maxwell son una generalización genial de leyes experimentales. Los resultados obtenidos con estas ecuaciones han confirmado ampliamente su validez, y su explicación rigurosa ha sido siempre muy difícil, si se recuerda que la definición de los campos presupone la existencia de cargas eléctricas o partículas magnéticas. Ha sido necesario recurrir a la teoría de la relatividad para encontrar una explicación plausible. Sin embargo, desde su origen han explicado satisfactoriamente los fenómenos de propagación de ondas electromagnéticas, incluidas las luminosas. Aquí las aceptaremos también con la única justificación que hemos dado, sin más explicaciones sobre su origen o demostración.

Tomando rotacionales en las ecuaciones tercera y cuarta y teniendo en cuenta las otras, puede eliminarse una de las variables, obteniendo para la otra las **ecuaciones de propagación** (una por cada variable), llamando al producto $\epsilon\mu = 1 / c^2$, (c es la velocidad de la luz, 300.000 Km./seg., como veremos después:

La primera ecuación vectorial se desdobla en tres ecuaciones, una para cada componente del vector E:

$$\nabla^2\vec{E}=\frac{1}{c^2}\frac{\partial^2\vec{E}}{\partial t^2}$$
$$\nabla^2\vec{H}=\frac{1}{c^2}\frac{\partial^2\vec{H}}{\partial t^2}$$

$$\frac{\partial^2 E_x}{\partial x^2}+\frac{\partial^2 E_x}{\partial y^2}+\frac{\partial^2 E_x}{\partial z^2}=\frac{1}{c^2}\frac{\partial^2 E_x}{\partial t^2}$$
$$\frac{\partial^2 E_y}{\partial x^2}+\frac{\partial^2 E_y}{\partial y^2}+\frac{\partial^2 E_y}{\partial z^2}=\frac{1}{c^2}\frac{\partial^2 E_y}{\partial t^2}$$
$$\frac{\partial^2 E_z}{\partial x^2}+\frac{\partial^2 E_z}{\partial y^2}+\frac{\partial^2 E_z}{\partial z^2}=\frac{1}{c^2}\frac{\partial^2 E_z}{\partial t^2}$$

La solución de estas ecuaciones debe satisfacer además las condiciones en los límites, que dependen de cada aplicación particular. Aquí vamos a estudiar brevemente dos tipos de soluciones que interesan especialmente en Electrónica:

La **propagación libre en forma de onda plana**, que es la forma más habitual de usar las ondas electromagnéticas en comunicaciones, radar y localización, etc...

Las **guías de onda**, que son canalizaciones por donde se orienta la propagación dentro de conductos metálicos, análogas a las líneas de transmisión.

14.14.- Propagación por onda plana.

Vamos a probar una solución con campo eléctrico vertical, (campo magnético perpendicular, horizontal) constante en un plano perpendicular al eje z, que es el de la dirección de propagación. La solución debe ser sinusoidal en función de z y t (debido al segundo orden de la ecuación), por lo que será de la forma:

$$E_x=0$$
$$E_y=K_y sen(\omega t-\beta z)$$
$$E_z=0$$

La constante de fase β expresa la variación de fase (retardo) con z. En un momento determinado, por ejemplo t=0, la onda es sinusoidal a lo largo del eje z, su valor se repite cada vez que βz aumenta 2π, es decir, cada vez que z aumenta en la magnitud **$\lambda = 2\pi/\beta$ llamada longitud de onda**.

Sustituyendo la expresión propuesta de E_y en la segunda ecuación del párrafo anterior, se obtiene la relación:

En un punto determinado, por ejemplo, z=0, la pulsación o

$$-K_y\beta^2 sen(\omega t-\beta z) = -K_y \frac{\omega^2}{c^2} sen(\omega t-\beta z)$$

$$\frac{\omega^2}{\beta^2}=c^2 \; ; \; \frac{\omega}{\beta}=c$$

$$2\pi \frac{f}{\frac{2\pi}{\lambda}}=f\lambda=c$$

variación de fase con el tiempo es ω.

La **velocidad de avance de la onda o velocidad de fase** se deduce fácilmente haciendo la fase constante y diferenciando:

$$\omega t-\beta z=cte. \; ; \; \omega dt-\beta dz=0$$

$$\frac{dz}{dt}=\frac{\omega}{\beta}=f\lambda=c=\frac{1}{\sqrt{\epsilon\mu}}$$

La velocidad de avance de la onda es pues una función de la constante dieléctrica y de la permeabilidad del medio y es independiente de la frecuencia. Como la luz es también una onda electromagnética, la velocidad de avance de las ondas es **c = 300.000 Km. / seg.**, para cualquier frecuencia. Recuérdese también la importantísima relación **$f\lambda$ = c**, aplicable también a cualquier frecuencia. La constante K_y , que define la amplitud de la onda, depende del generador.

En el capítulo de propagación y antenas utilizaremos este tipo de propagación "libre".

14.15.- Guías de onda.

Las guías de onda transmiten las señales por medio de ondas electromagnéticas canalizadas. Las guías de onda son conductos huecos de paredes conductoras. Dentro de estos conductos se "inicia" una onda electromagnética que se refleja en las paredes y se propaga a lo largo del conducto sin sufrir otra atenuación que la debida a la resistividad no nula de las paredes.

Debido a las limitaciones que veremos a continuación, las guías de onda sólo pueden utilizarse con longitudes de onda comparables con su dimensión, lo que limita su uso a frecuencias muy altas, llamadas **microondas**, del orden de 1.500 MHz. en adelante, para que sus dimensiones sean razonables.

La propagación obedece a las ecuaciones de Maxwell y debe, además, satisfacer las condiciones en los límites, en las paredes conductoras. Estas condiciones son:

El campo eléctrico debe ser perpendicular a las paredes (o nulo en ellas).

El campo magnético, perpendicular al eléctrico, debe ser

paralelo a las paredes o nulo en ellas.

14.16.- Solución de las ecuaciones para las guías de onda.

Si consideramos la guía de onda de sección rectangular, de dimensiones a y b, representada en la figura 14-10, tomando las coordenadas indicadas, con la dirección z en el sentido de la conducción, perpendicular al papel, podemos probar la solución:

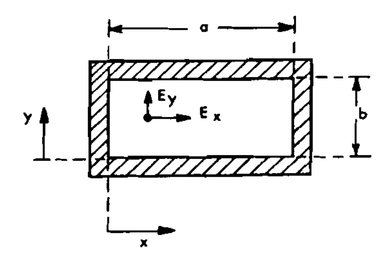


Fig. 14-10.

$$E_y = K_y sen\frac{n\pi x}{a}\cos\frac{m\pi y}{b} sen(\omega t - \beta z)$$
$$E_x = 0 \qquad E_z = 0$$

preparada para que E_y se anule para x=0 y x=a (o sea, en las paredes) y pase por un máximo (perpendicular a las paredes) cuando y=0 e y=b.

Esta solución es una ecuación de propagación o avance de la onda a lo largo del eje z, a una velocidad dada por la relación $\omega t - \beta z$ = constante, que, diferenciada, nos da la velocidad de propagación, ya que $\omega dt - \beta dz = 0$, es decir, la velocidad de propagación de la onda es $v = dz / dt = \omega/\beta$.

La constante β, constante de fase, hace que se repita el valor de la amplitud cada vez que βz aumente 2π, es decir cada vez que z aumente el valor $2\pi/\beta$ llamado λ, **longitud de onda.** La velocidad de avance es pues $\omega/\beta = 2\pi f/\beta$, o sea, $\mathbf{v = f\lambda}$.

Esta solución corresponde al llamado **modo transversal n,m** porque el campo eléctrico es transversal ($E_Z = 0$). n y m son los números de máximos del campo eléctrico con la variación de x e y entre 0 y el máximo (de una pared a otra). El campo eléctrico tiene la dirección del eje y, el campo magnético será en la dirección x.

Los parámetros que figuran en la solución no son independientes, sino que deben satisfacer a la ecuación de E_y del párrafo anterior (es evidente que las otras dos ecuaciones se satisfacen para $E_x = 0$ y $E_z = 0$). Para ver la relación entre ellos, calcularemos las derivadas parciales segundas, que son:

$$\frac{\partial^2 E_y}{\partial x^2}=-\frac{n^2\pi^2}{a^2}K_y sen\frac{n\pi x}{a}\cos\frac{m\pi y}{b}sen(\omega t-\beta z)=-\frac{n^2\pi^2}{a^2}E_y$$
$$\frac{\partial^2 E_y}{\partial y^2}=-\frac{m^2\pi^2}{b^2}K_y sen\frac{n\pi x}{a}\cos\frac{m\pi y}{b}sen(\omega t-\beta z)=-\frac{m^2\pi^2}{b^2}E_y$$
$$\frac{\partial^2 E_y}{\partial z^2}=-\beta^2 K_y sen\frac{n\pi x}{a}\cos\frac{m\pi y}{b}sen(\omega t-\beta z)=-\beta^2 E_y$$
$$\frac{\partial^2 E_y}{\partial t^2}=-\omega^2 K_y sen\frac{n\pi x}{a}\cos\frac{m\pi y}{b}sen(\omega t-\beta z)=-\omega^2 E_y$$

Sustituyendo estos valores en la ecuación de E_y :

$$-\frac{n^2\pi^2}{a^2}E_y-\frac{m^2\pi^2}{b^2}E_y-\beta^2 E_y=-\frac{\omega^2}{c^2}E_y$$

de donde se puede obtener el valor de β :

$$\beta^2=\frac{\omega^2}{c^2}-\frac{n^2\pi^2}{a^2}-\frac{m^2\pi^2}{b^2}=\frac{4\pi^2}{\lambda^2}-\frac{n^2\pi^2}{a^2}-\frac{m^2\pi^2}{b^2}$$

En esta expresión, n y m pueden ser cualesquiera números que definen el **modo**. Sin embargo su valor está limitado para que β no se haga negativa.

En la práctica, los valores de a y b se diseñan de forma que sólo sea posible el **modo 1,0**, es decir n=1 y m=0. Para este modo, las ecuaciones se convierten en :

$$E_x=0$$
$$E_y=Ksen\frac{\pi x}{a}sen(\omega t-\beta z)$$
$$\beta^2=\frac{4\pi^2}{\lambda^2}-\frac{\pi^2}{a^2}$$
$$\beta=2\pi\sqrt{\frac{1}{\lambda^2}-\frac{1}{4a^2}}$$

El valor máximo de λ que se puede propagar por la guía, llamado **longitud de onda de corte** es 2a, la mínima frecuencia, llamada **frecuencia de corte** es **c/2a**.

La **longitud de onda de la línea** es:

$$\lambda_g=\frac{2\pi}{\beta}=\frac{1}{\sqrt{\frac{1}{\lambda^2}-\frac{1}{4a^2}}}=\frac{\lambda}{\sqrt{1-\left(\frac{\lambda}{2a}\right)^2}}$$

superior a la del espacio libre (vacío o aire). La velocidad de propagación a lo largo de la guía es $f\lambda_g > f\lambda = c$, **mayor que la velocidad de la luz.**

Para modos superiores al 1,0, la longitud de onda de corte (obtenida de la expresión de β) es:

$$\lambda_c=\frac{2}{\sqrt{\left(\frac{m}{b}\right)^2+\left(\frac{n}{a}\right)^2}}$$

inferior a la del modo principal. Las guías de onda se proyectan de forma que la frecuencia de uso sea ligeramente superior a la de corte del modo 1,0, pero sea inferior a la de los demás modos, de forma que sólo se pueda propagar el modo 1,0, principal.

14.17. Imagen física de la propagación.

La solución estudiada corresponde a una propagación longitudinal (por estar t y z en el mismo paréntesis, directamente relacionadas) con un modo transversal de distribución del campo eléctrico vertical que pasa por un máximo en el centro y se anula en los bordes. Las líneas de fuerza de este campo se pueden representar como se ve en la figura 14-11, donde la amplitud del campo está representada por líneas más juntas, de trazo continuo. Las líneas del campo magnético horizontal asociado se cierran sobre sí mismas como se ve en la figura, satisfaciendo las ecuaciones de Maxwell. En un momento dado, el campo eléctrico es máximo en una sección determinada, la del dibujo, y varía a lo largo de la guía, anulándose después de un cuarto de longitud de onda e invirtiéndose a la media longitud de onda.

La relación entre el campo eléctrico y el magnético es constante y, como puede calcularse a partir de las ecuaciones de Maxwell, vale:

$$Z_o=\frac{377\lambda_g}{\lambda}\ ohmios$$

Esta magnitud es la equivalente de la impedancia característica de las líneas de transmisión, en una analogía entre las guías y las líneas en que los campos eléctricos y magnéticos son los análogos de la tensión e intensidad. La onda

descrita hasta aquí es la análoga a la onda directa, viaja en una dirección, desde el generador a la carga.

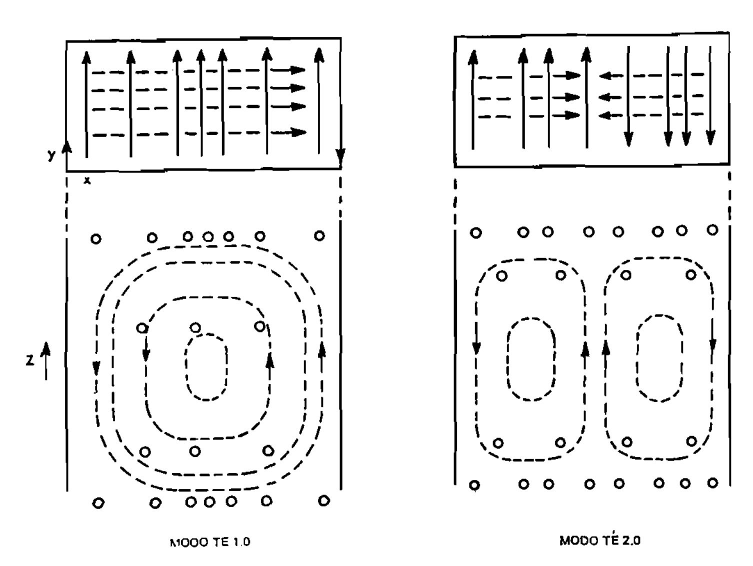


Fig. 14-11.

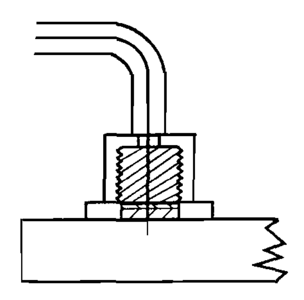


Fig. 14-12.

El generador es un transductor consistente en una pequeña antena que inicia la onda electromagnética y puede ser el extremo de una línea coaxial, como se representa en la figura 14-12.

En la carga, la línea puede terminar en otra antena o transductor coaxial. Si la terminación no tiene la impedancia característica, se producen reflexiones, como en las líneas, y también ondas estacionarias a lo largo del eje z.

Las reflexiones en las paredes laterales ya han sido tenidas en cuenta al suponer campo eléctrico nulo en ellas, como si lateralmente las ondas tuviesen una terminación en cortocircuito y se produjesen ondas estacionarias transversales. Este efecto da lugar a una imagen atractiva de la propagación representada en la figura 14-13, que descompone la onda que avanza en dos ondas cruzadas que interfieren entre sí en el sentido

transversal, dando ondas estacionarias en esa dirección y se suman en sentido longitudinal produciendo la onda directa de propagación según ese eje.

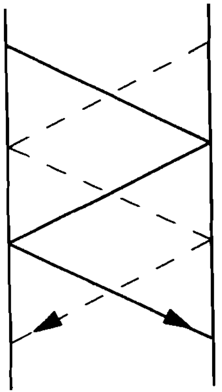


Fig. 14-13.

14.18.- Guía en cuarto de onda. Cavidades resonantes.

Si una guía se termina en un cortocircuito, se produce una reflexión total del campo eléctrico y un onda estacionaria en el sentido longitudinal, análoga a la que se producía en las líneas de transmisión. A un cuarto de onda del final (o cuarto más un número entero de longitudes de onda) se refleja una alta impedancia y la guía "no carga" en ese punto.

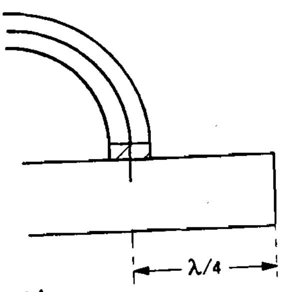


Fig. 14-14.

Utilizando esta propiedad existen multitud de dispositivos, como terminaciones después de un conector coaxial (fig. 14-14) que tratan de dar una terminación aproximada en el conector y nula absorción por parte de la guía que sigue (la energía que no es recibida por el conector en el camino hacia la derecha, es reflejada y devuelta por la guía en cuarto de onda).

Es preciso señalar que en guías de onda no existe un análogo a la línea abierta, por no disponer de eficaces cortocircuitos magnéticos.

Otra aplicación de las reflexiones en un final cortocircuitado es la **cavidad resonante**, que es una línea terminada en cortocircuito a ambos lados, formando una caja semicerrada de longitud media longitud de onda de esa línea.

Además de las ondas estacionarias **laterales** se forman ondas estacionarias **longitudinales**, con un máximo en el centro, que equivale a una **resonancia**. La alimentación puede hacerse a través de "ventanas", que ven una impedancia moderada, multiplicándose la onda de llegada para formar un fuerte máximo en el centro. Un ejemplo de aplicación es la **cavidad TR** del radar, que activa un dispositivo de ionización en el centro que cortocircuita la cavidad cuando las señales son muy grandes, protegiendo así al circuito receptor de grandes sobretensiones.

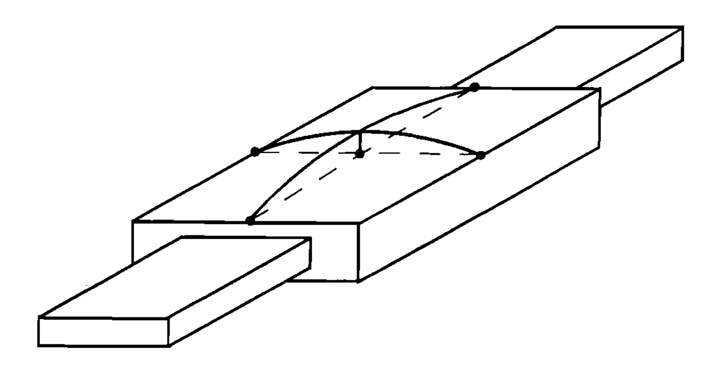


Fig. 14-15.

14.19.- Componentes de guía de onda.

Las guías de onda se utilizan en fragmentos rectos que se acoplan unos a otros con "bridas", proyectadas cuidadosamente para que haya una continuidad eléctrica en las paredes. También pueden curvarse ligeramente y formar "codos", guías flexibles, etc... También existen componentes que equivalen a constantes concentradas, como las "ventanas" de la figura 14-16. Una ventana como la de la izquierda de la figura acorta el recorrido del campo eléctrico, equivaliendo a un "condensador". La ventana de la derecha concentra el campo magnético, es "inductiva".

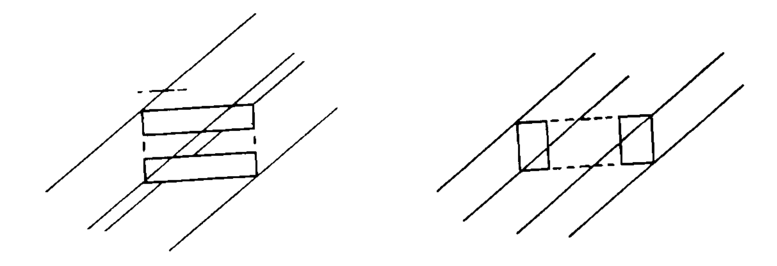


Fig. 14-16.

Normalmente, este tipo de componentes y otros tales como tornillos verticales, etc... se utilizan para "adaptar"

experimentalmente las cargas y evitar reflexiones.

Otra componente típica es la **carga artificial**, que va cargando gradualmente (para evitar reflexiones) el campo eléctrico para absorber la energía incidente. Se usa para el mantenimiento de equipos, etc...(fig 14-17).

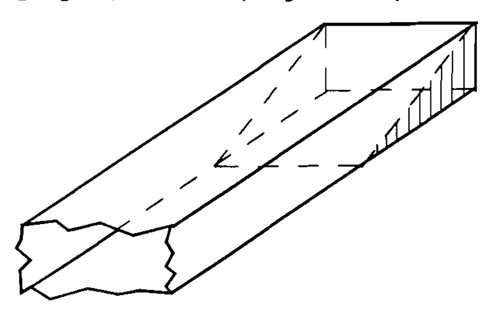


Fig. 14-17

Un dispositivo interesante es el **acoplamiento direccional** de la figura 14-18. Está constituido por dos guías de onda acopladas con dos orificios. Una de ellas es la "guía principal" y la otra es la "acoplada", que tiene dos salidas, como se ve en la figura. Los dos orificios dejan pasar a la guía acoplada una pequeñísima parte de la energía que viaja por la principal. Los orificios están alejados una longitud de un cuarto de longitud de onda más un múltiplo entero de media longitud de onda.

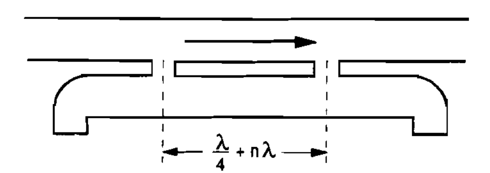


Fig. 14-18.

La onda que viaja de izquierda a derecha, pasa (multiplicada por un factor constante) a través del primer orificio y se propaga en la acoplada en ambas direcciones, a partir de ese punto. En el segundo orificio ocurre lo mismo. La onda que viaja hacia la derecha, desde el primer orificio, se encuentra **en fase** con la acoplada por el segundo (ambas tienen el mismo trayecto) y que también viaja hacia la derecha, con lo cual ambas ondas se refuerzan y salen por la derecha (reducidas en el factor de acoplamiento). En cambio, la onda que pasa por el segundo orificio y retrocede hacia el primero se encuentra en contrafase con la que entra por el primero, porque su tiene que recorrer dos veces la distancia entre orificios y su trayecto difiere de la del primero en media longitud de onda más un múltiplo entero de

la longitud de onda. Ambas ondas se cancelan y no sale ninguna señal apreciable por la izquierda.

Como el mismo razonamiento podría hacerse hacia la izquierda, se deduce que las ondas que **viajan hacia la derecha salen por la derecha y las que viajan hacia la izquierda, por la izquierda**, reducidas ambas en el factor de acoplamiento. Esta propiedad puede usarse para **medir** la onda incidente (hacia la derecha) y medir o minimizar las reflexiones, etc...

CAPITULO 15

ANTENAS Y DIAGRAMAS DE RADIACION

15.1.- Introducción.

En el capítulo anterior hemos estudiado las nociones de propagación de ondas electromagnéticas que hemos considerado necesarias para su aplicación a los sistemas electrónicos. No se ha pretendido entrar en la teoría del electromagnetismo, de la que se supone que el lector tiene algunos conocimientos, por considerar que no es el propósito de este libro tratar con más profundidad esta teoría.

En este capítulo estudiaremos las antenas como transductores de señales eléctricas a electromagnéticas, con el nivel mínimo necesario para su uso en los sistemas electrónicos y tampoco entraremos en la demostración de algunas de sus propiedades, que pueden encontrarse en la literatura especializada. También estudiaremos los **diagramas de radiación** de las antenas o sistemas de antenas, que, como veremos, son una herramienta indispensable para "marcar" el espacio, realmente la única herramienta de que disponemos para ello.

15.2.- Características de las antenas.

Las antenas son transductores que convierten las señales eléctricas en electromagnéticas **radiando** energía electromagnética en el espacio libre, generalmente vacío o aire, o en el medio en que están situadas. La radiación obedece a las leyes del electromagnetismo. La propagación ocurre satisfaciendo las ecuaciones de Maxwell y las condiciones en los límites marcadas por las leyes de Ampère y Lentz. La **velocidad** de propagación de las ondas electromagnéticas es la de la luz, **c = 300.000 Km./seg.**, y la forma de onda es sinusoidal, como vimos en el capítulo anterior. Despreciando el tamaño de la antena a partir de cierta distancia, la antena puede considerarse puntual, centro de la propagación esférica, que considerada a mucha distancia puede considerarse como **onda plana**.

La propagación ocurre a la misma velocidad en todas direcciones, pero no ocurre con la misma amplitud en todas direcciones, sino que da lugar a un **diagrama de radiación**, que es la representación del campo eléctrico a distancia constante (grande con relación a la longitud de onda y al tamaño de la antena) en función de la posición angular en coordenadas polares, por ejemplo amplitud en función del acimut y elevación. El corte de ese diagrama espacial con el plano horizontal da lugar al **diagrama horizontal**, de amplitud en función del acimut, y el corte con el plano vertical da lugar al **diagrama vertical**, de amplitud en función del ángulo de elevación. Generalmente en los diagramas de radiación se representan valores relativos, con lo cual no es necesario especificar la distancia a la que se hace la medida.

La potencia radiada en una dirección por unidad de superficie es proporcional al producto de la amplitud del campo eléctrico por la del magnético. El vector que representa la potencia de la onda por unidad de superficie se llama **vector Poynting**. Como la relación entre el campo eléctrico y el magnético es constante e igual a la impedancia característica, el diagrama de radiación en **potencia**, que representa la amplitud del vector Poynting en polares, es el **cuadrado del diagrama en amplitud**, salvo factores constantes.

La **ganancia** de la antena en una dirección determinada es la potencia radiada por unidad de superficie en esa dirección dividida por el valor que correspondería si la antena radiase uniformemente en todas direcciones. La ganancia puede también definirse en **amplitud** y la ganancia en potencia es el cuadrado de la ganancia en amplitud.

La ganancia en la dirección en la que es máxima se llama **directividad**. En algunos libros se llama también ganancia al producto de la directividad por el rendimiento de la antena, para tener en cuenta las pérdidas de radiación, aunque esta denominación es algo confusa por lo que no la utilizaremos aquí. La potencia media radiada por unidad de superficie a una distancia R será $W_m=\frac{W}{4\pi R^2}$ siendo W la potencia total. La representación del valor de amplitud del vector Poynting dividida por W_m es el diagrama de radiación en ganancia.

15.3.- Dipolo infinitesimal.

Si consideramos una antena formada por un fragmento infinitesimal de conductor llamado **dipolo** alimentado por una corriente sinusoidal de amplitud I, el diagrama de radiación será simétrico respecto a la recta que contiene al dipolo. El campo eléctrico y el magnético en función de la inclinación respecto al dipolo, de la distancia r, grande respecto a la longitud de onda, y de la longitud l (pequeña) del dipolo, puede demostrarse que valen:

$$E=\frac{Il}{4\pi r}sen\theta[\omega\mu\cos\omega(t-\frac{r}{c})]$$
$$H=\frac{Il}{4\pi r}sen\theta[\frac{\omega}{c}\cos\omega(t-\frac{r}{c})]$$

A distancias pequeñas, además de este campo hay otro inducido que es inversamente proporcional a r y puede despreciarse a distancias grandes comparadas con la longitud de onda.

Los campos son perpendiculares entre sí y perpendiculares al radio que une el punto con la antena. Si el dipolo es vertical, el campo eléctrico en un punto está en el plano vertical que contiene el dipolo y el punto, perpendicular al "radio" o línea que une al punto con la antena. El campo magnético es perpendicular a ese plano en ese punto. La relación entre las amplitudes de ambos campos es la impedancia característica del aire que vale (unidades eléctricas):

$$Z=\frac{E}{H}=c\mu=\frac{1}{\sqrt{\epsilon\mu}}\mu=\sqrt{\frac{\mu}{\epsilon}}=377\,ohmios.$$

La figura 15-1 representa el diagrama vertical en amplitud, la 15-2 el diagrama en potencia (cuadrado del anterior).

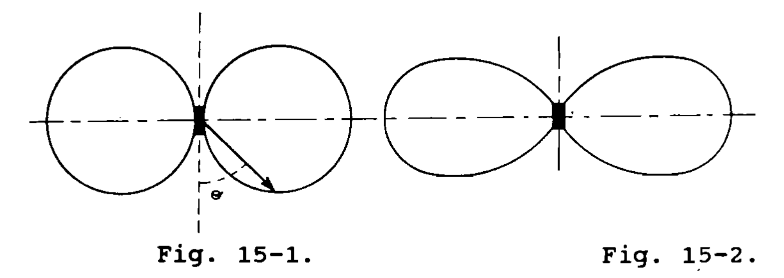


Fig. 15-1. Fig. 15-2.

La ganancia en el plano ecuatorial vale 1,5 (en potencia).

15.4.- Resistencia de radiación.

Se define como resistencia de radiación al valor de la potencia radiada dividida por el cuadrado de la intensidad, por analogía con la resistencia eléctrica, o sea :

$$R_{rad.}=\frac{P_{rad.}}{I^2}$$

En el caso del dipolo infinitesimal vale:

$$R_{rad.} = \frac{20\omega^2 l^2}{c^2} \, ohmios$$

$$R_{rad.} = \frac{20\omega^2 l^2}{c^2} \, ohmios$$

15.5.- Dipolo de media onda.

Si en lugar de un dipolo elemental se toma un radiador constituido por dos fragmentos de un cuarto de longitud de onda que forman un **dipolo de media longitud de onda** se obtiene una distribución de intensidad como la representada en la figura 15-3, anulándose en los extremos y con un máximo en el centro, como si la antena fuese una línea abierta de cuarto de longitud de onda y se formase una onda estacionaria con un máximo en el centro.

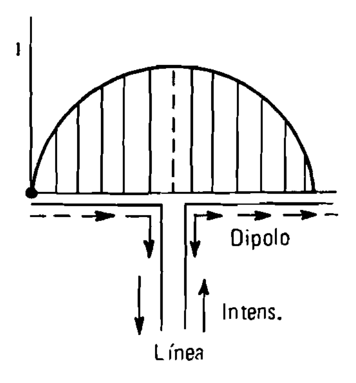


Fig. 15-3.

El diagrama de radiación en amplitud (campo eléctrico) puede demostrarse que es :

$$E_{\theta,r} = \frac{2I}{cr} \frac{\cos(\pi/2\cos\theta)}{sen\theta} \cos\omega\left(t - \frac{r}{c}\right)$$

El diagrama, representado en la figura 15-4, es algo más directivo que el del dipolo infinitesimal. La ganancia máxima es 1,64 y la resistencia de radiación es 73,5 ohmios.

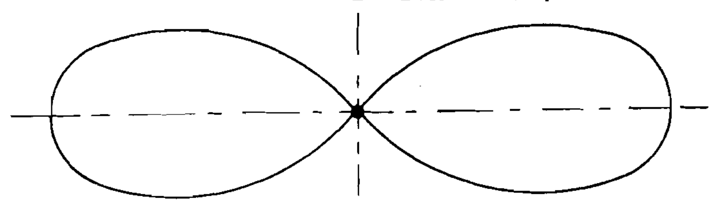


Fig. 15-4.

15.6.- Radiadores de más de media longitud de onda.

Cuando la longitud de la antena es superior, en el plano de la antena se obtienen diagramas más complicados, con lóbulos en número creciente con la longitud de la antena. El diagrama de radiación depende también de la alimentación de la antena, que determina la distribución de intensidad a lo largo de la misma. Si se alimenta como en la figura 15-3 en un máximo de onda estacionaria y la longitud es múltiplo de media longitud de onda, la intensidad a lo largo de la antena se distribuye como una onda estacionaria que se anula en los extremos, pero formando varios lóbulos con nodos intermedios, de forma que la intensidad cambia de signo al pasar cada nodo. El diagrama de radiación se obtiene integrando el correspondiente a cada fragmento elemental, o dividiendo la antena en fragmentos cortos conocidos, resultando diagramas como los indicados en la figura 15-5. En la práctica no se usan antenas de más de media longitud de onda, por lo que no las trataremos con más detalle.

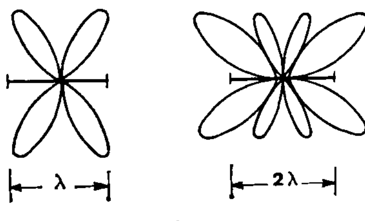


Fig. 15-5.

15.7.- Sistemas de antenas.

Generalmente se obtienen diagramas de radiación más complicados agrupando antenas sencillas, omnidireccionales o directivas, para que la interferencia entre las radiaciones genere el diagrama de radiación deseado. Lo más frecuente es que se usen antenas omnidireccionales en estos sistemas, pero a veces se usan muchas antenas iguales y directivas, multiplicándose entonces el diagrama de radiación del sistema, obtenido como si las antenas fuesen omnidireccionales, por el diagrama individual de cada antena.

Como norma general, el campo eléctrico radiado por un sistema **a gran distancia del sistema** se obtiene sumando los campos producidos por cada una de las antenas para cada dirección. Si se usa una representación compleja del campo, la suma debe hacerse a una distancia normalizada, teniendo en cuenta la amplitud y fase de los diferentes campos parciales. Si se supone que la distancia **r** es grande, los campos parciales tienen la misma dirección, perpendicular al "radio" o recta que une el punto donde se mide el campo con el **centro de fase** del sistema. El centro de fase es un punto de referencia arbitrario, generalmente el centro del sistema si lo hay, que sirve de referencia o fase cero para la alimentación en ese punto. Las demás rectas que unen el punto de medida del campo con las diferentes antenas del sistema se consideran todas paralelas y las antenas mismas **puntuales**, sustituidas por un punto de

referencia cada una de ellas que es en realidad su centro de fase en el que se considera la fase inicial (de la alimentación) respecto a la del centro del sistema. En el párrafo siguiente se ilustrarán mejor estos conceptos a la vista de un ejemplo concreto.

15.8.- Diagrama de radiación de un par de antenas iguales omnidireccionales.

Sean las antenas iguales A_1 y A_2 distanciadas **d radianes (es decir d es la distancia multiplicada por 2**π/λ) y alimentadas por corrientes de igual amplitud, pero fases diferentes. Suponemos las antenas omnidireccionales en el plano horizontal, que es el representado en la figura 15-6. Tomemos como centro de fase el centro del segmento que las une. Las antenas se suponen puntuales o son antenas verticales cuyos centros son los puntos representados. Las fases relativas a ese centro virtual sean φ y $-\varphi$, lo que comprende cualquier combinación posible, ya que sólo expresa que se ha tomado como referencia de fase el valor medio de las fases de las dos antenas. Por ejemplo, si están en fase sus fases serán ambas 0; si están en contrafase, se tomarán como de fases $\pi/2$ y $-\pi/2$, etc...

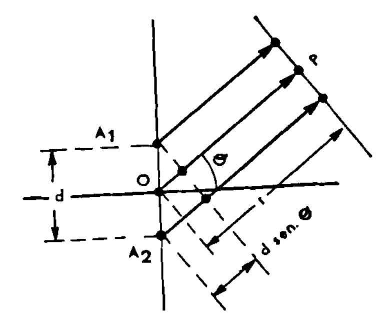


Fig. 15-6.

Para un punto P situado a una distancia **r** grande, el camino a recorrer por la onda procedente de la antena A_1 es r menos la magnitud d/2 sen θ y para la onda procedente de A_2 es r más la magnitud d/2 sen θ. La distancia r también en radianes eléctricos, es decir, es la distancia multiplicada por $2\pi/\lambda$.

El campo en P será pues la suma de dos campos de igual amplitud (sea ésta A) y de fases $\varphi+d/2 \mathrm{sen}\theta-r$ y $-\varphi-d/2 \mathrm{sen}\theta-r$. Si sólo nos interesa la amplitud del campo resultante, será la suma de dos vectores de desfase $2\varphi+d \mathrm{sen}\theta$ y módulo A. (También se podría aplicar el mismo procedimiento si su amplitud no fuese igual). La suma vale:

$$2A \cos\ [\varphi+\frac{d}{2} sen\theta]$$

y esta expresión es también la del diagrama de radiación. Podría también haberse obtenido gráficamente dibujando en cada dirección los dos vectores y sumándolos. Por ejemplo, si $\varphi=0$, antenas en fase, y $d=\lambda/2$, antenas espaciadas media longitud de onda, el ángulo de desfase es $\pi\text{sen}\theta$ y vale cero para $\theta=0$ o múltiplo de π, dando un máximo, y vale π para $\theta=\pi/2, 3\pi/2$, etc... dando cero en esas direcciones. El diagrama está representado en la figura 15-7.

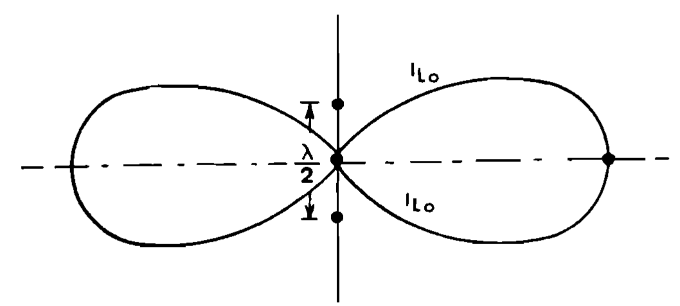


Fig. 15-7.

Los cálculos anteriores pueden hacerse para varias distancias y varios desfases, obteniendo los diagramas de la figura 15-8.

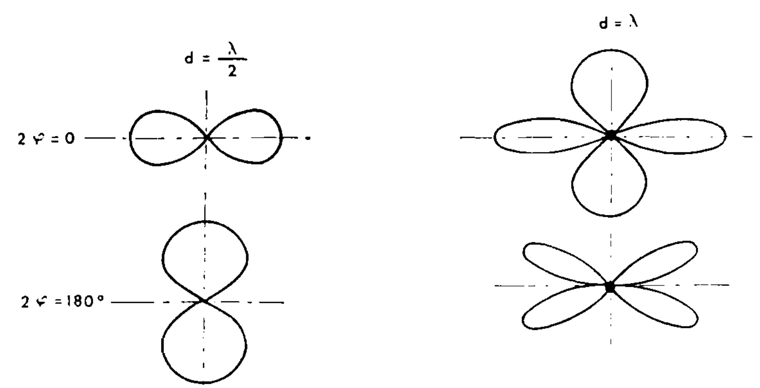


Fig. 15-8.

En todos estos casos se ha calculado el diagrama horizontal del sistema, suponiendo las antenas verticales y omnidireccionales. Si las dos antenas no fuesen omnidireccionales, sino que tuviesen un diagrama propio, se multiplicaría el diagrama obtenido por el de las antenas (que deben ser iguales). En el plano vertical también se multiplica cada valor horizontal (en el acimut que corresponda) por el vertical común a las antenas, obteniéndose así el diagrama en tres dimensiones.

15.9.- Reflexiones. Antenas imagen.

Las ondas electromagnéticas se propagan a la velocidad de

la luz. La constante c aparece en las ecuaciones de Maxwell como función de la constante dieléctrica y la permeabilidad magnética del medio. Al cambiar de medio, la velocidad en el nuevo medio es distinta de la del aire o vacío y se producen reflexiones y refracciones como las estudiadas para la luz en óptica.

En particular, si la onda electromagnética se encuentra una superficie de discontinuidad de las dos constantes citadas, con un nuevo valor de la impedancia característica, la onda se comporta como la de una línea de transmisión terminada en una impedancia distinta de la característica. El caso más interesante es cuando se encuentra con una superficie conductora, como el suelo húmedo, que produce una reflexión total que anula la componente del campo eléctrico paralela al suelo.

La figura 15-9 representa las condiciones de reflexión para una onda incidente con una dirección de polarización cualquiera, lo que no es habitual. La componente horizontal del campo eléctrico se invierte al reflejarse, mientras que la componente vertical se refleja sin cambios. La figura 15-10 representa el caso más habitual de onda incidente con polarización perpendicular a la dirección de propagación, y es un caso particular de la anterior.

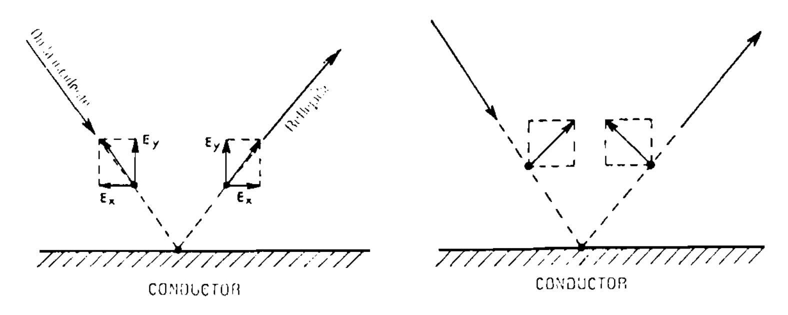


Fig. 15-9. Fig. 15-10.

Del mismo modo que ocurría en las guías de onda, la reflexión da lugar a interferencias entre la onda incidente y la reflejada, algunas de las cuales estudiaremos a continuación.

Del mismo modo que en óptica, la reflexión ocurre **como si la onda reflejada procediese de una antena imagen**, cuya fase debe cumplir las condiciones en los límites. En el caso de la antena vertical, la antena imagen es simétrica de la real, pero no así la fase. Ambas antenas, real e imaginaria están **en fase**. Como se ve en la figura 15-11, al punto P llegan la onda directa y la reflejada, que es como si procediese de la imagen. La interferencia entre estas dos ondas forma el **diagrama de radiación** que veremos después.

En el caso de antena horizontal, la onda reflejada tiene fase opuesta a la directa. Esto puede considerarse, como se ve en la figura 15-12, como si la **antena imagen** tuviese una fase opuesta a la real. Es decir, hay una antena imagen, simétrica de

la real, con una fuese que tampoco esta vez es la simétrica de la real, sino la opuesta.

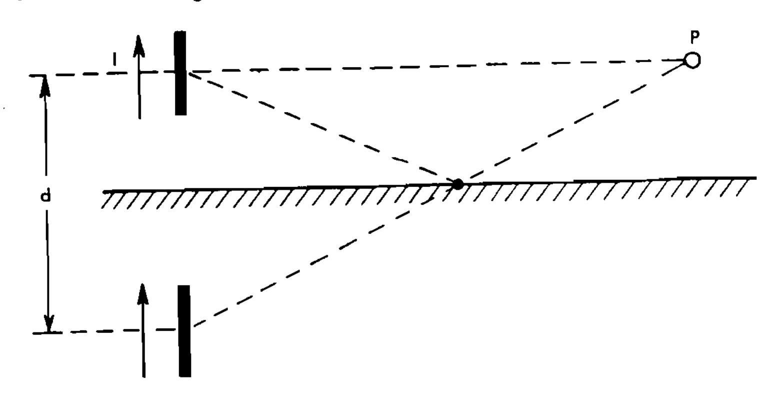


Fig. 15-11.

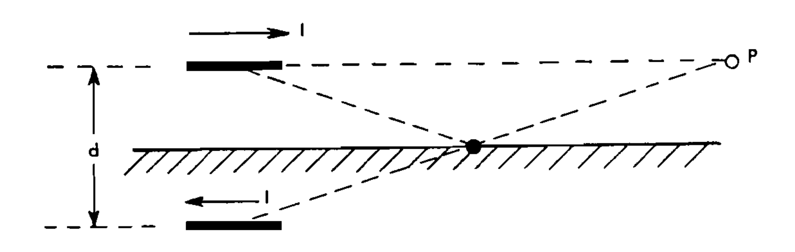


Fig. 15-12.

La tierra refleja bastante, aunque no sea un conductor ideal. La reflexión depende de la naturaleza y humedad del terreno. Cuando se quiere que estas reflexiones estén bien determinadas, se pone una red metálica conductora en las proximidades de la antena. Esto se hace, por ejemplo, en sistemas de localización aeronáutica como el ILS o el VOR, donde interesa que el diagrama de radiación esté bien determinado. En otros casos, para evitar las reflexiones se usan antenas directivas que radian poco hacia la tierra, como en el caso del radar, o se alejan de tierra las antenas, para que la energía de la onda reflejada sea despreciable frente a la directa.

15.10.- Reflectores parabólicos. Plano radiante.

El modo más sencillo de concentrar la radiación procedente de un punto (o antena pequeña) en una dirección determinada es el reflector parabólico, debido a la propiedad de que todas las ondas procedentes del foco de la parábola salen en la misma dirección paralela al eje de la parábola. Cuanto más grande es el reflector parabólico, mayor ángulo sólido procedente del foco será reflejado en esa dirección. El diagrama de radiación puede

obtenerse teniendo en cuenta la propiedad de la parábola de que es el lugar geométrico de los puntos equidistantes del foco y de la directriz, o, en el caso de reflector parabólico de revolución, los puntos de la antena son equidistantes del foco y del plano director. La figura 15-13 ilustra esta propiedad. Los rayos reflejados en la antena parabólica pueden suponerse procedentes del plano director, que actúa así como antena imagen.

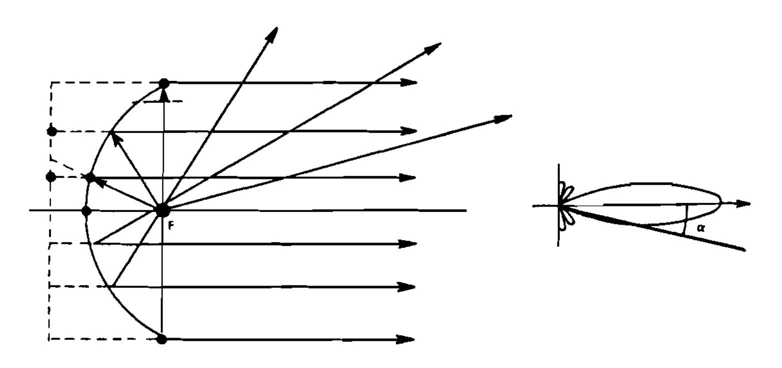


Fig. 15-13.

El diagrama de radiación de una antena parabólica "iluminada" en su foco es pues el de **un plano radiante** que sustituye al sistema foco-reflector. Cuando el reflector se corta de forma rectangular, se obtiene una **apertura rectangular radiante**.

Se puede demostrar que este diagrama de radiación (en el plano horizontal) resulta ser la **función de muestreo**. Si se representa esta función en coordenadas polares, se obtiene un diagrama lobulado como el representado en la figura 15-13. El ancho del lóbulo principal (ángulo entre puntos a - 3 dB del máximo) resulta ser:

$$2\alpha = \frac{\lambda}{a}$$

siendo a el ancho de la apertura (dimensión horizontal del rectángulo). Por ejemplo, para una longitud de onda de 3 cm. y un ancho de antena de 3 m., el ancho del haz sería 0,01 radianes, aproximadamente 0,6 grados. Puede decirse que **este tipo de antenas es muy directivo**.

La ganancia en potencia resulta ser aproximadamente:

$$G = \frac{4\pi A}{\lambda^2}$$

15-11.- Antenas particulares.

En comunicaciones, las antenas más usadas son las verticales **de cuarto de onda**, que, junto con la antena imagen forman un dipolo que radia omnidireccionalmente en un plano horizontal. La figura 15-14 (a la izquierda) representa la distribución de corriente en la antena. La longitud de un cuarto de longitud de onda es a veces difícil de alcanzar y, por ello, se sustituye por una antena horizontal y su línea, por la cual circula un valor mayor de la intensidad (parte derecha de la figura). El área rayada representa la distribución de intensidad en ese caso. Las antenas receptoras suelen ser muy inferiores a un cuarto de longitud de onda. Como veremos, para el receptor interesa menos recibir mucha señal que recibir una mejor relación señal/ruido. Como la antena no es resonante, la relación señal/ruido es casi independiente de su longitud.

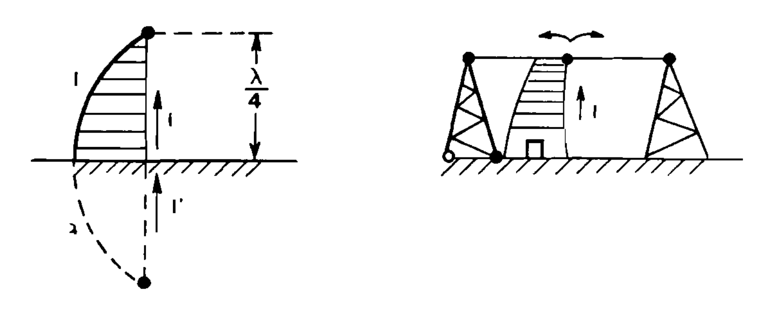


Fig. 15-14.

A partir de cierta frecuencia (VHF) se utilizan más las antenas directivas, especialmente los **dipolos** de media longitud de onda (por ejemplo, en televisión) con **reflectores** que dan, junto con la antena principal, el diagrama de radiación deseado.

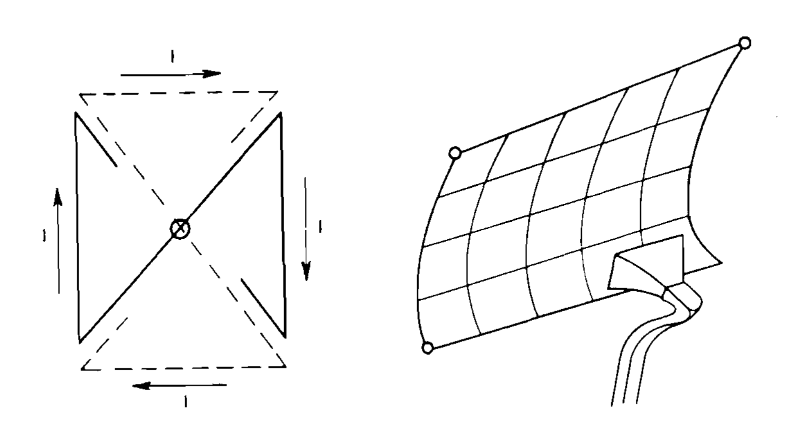


Fig. 15-15. Fig.15-16.

En aplicaciones aeronáuticas se utiliza mucho la antena omnidireccional de polarización horizontal llamada **Alford**, que consiste en dos dipolos doblados como los representados en la figura 15-15.

En radar, lo más frecuente es el reflector parabólico alimentado por una **bocina** de guía de onda. La bocina es una antena en la que termina la guía de onda e ilumina a su vez al reflector parabólico, como se ve en la figura 15-16.

La bocina sirve de transición entre la impedancia característica de la guía de onda y la del aire. Recuérdese que la impedancia característica de una guía de onda está relacionada con la longitud de onda en la guía y la longitud de onda en el aire a través de la expresión:

$$Z_g = \frac{377\lambda_g}{\lambda}$$

donde, a su vez, la longitud de onda en la guía es función del ancho de ésta a través de la expresión:

$$\lambda_g = \frac{\lambda}{\sqrt{1-(\lambda/2a)^2}}$$

Si se cambia a gradualmente hasta una apertura muy grande, se pasa de la impedancia característica de la guía al valor en el aire, 377 ohmios.

CAPITULO 16

MODULACION Y COMUNICACIONES

16.1.- Introducción.

Las señales contienen información a través de la codificación. La transmisión de información a larga distancia se hace por radiación de ondas electromagnéticas de alta frecuencia, que se propagan por el espacio. La transmisión por líneas se hace a veces también por impulsos de muy corta duración, cuyo espectro es también de alta frecuencia. Estas ondas o impulsos contienen la información a través de un tipo de codificación especial llamado **modulación**. La onda de alta frecuencia se llama **portadora**, modulada por alguno de los sistemas que veremos en este capítulo para transportar la información **moduladora**.

Las **comunicaciones** se hacen por medio de ondas moduladas. El **transmisor** realiza las funciones de modular una portadora y, una vez modulada, radiarla por una antena o enviarla por una línea de transmisión o guía de onda. El **receptor** recibe e identifica la transmisión (por la portadora) y deshace la modulación con circuitos llamados **detectores** para recuperar la moduladora, que es directa o indirectamente la información. A veces cambia previamente la frecuencia de la portadora para hacer la manipulación más fácil, en un proceso llamado **conversión**.

En todo el capítulo hablaremos de frecuencias llamando así a las frecuencias f en Hz. o a las pulsaciones ω en rad./s. Recuérdese que el cambio de frecuencia a pulsación se hace simplemente multiplicando por 2π. ($\omega = 2\pi f$).

16.2.- Producto de dos ondas sinusoidales. Espectro.

El espectro de una onda cosinusoidal de amplitud A_1 y pulsación ω_1 es el punto ω_1, A_1 de la figura 16-1. Del mismo modo, el espectro de la onda $A_2 \cos \omega_2$ es el punto ω_2, A_2. Si multiplicamos estas dos ondas y convertimos el producto de funciones trigonométricas en suma obtenemos:

$$P(t) = A_1 A_2 \cos\omega_1 t \cos\omega_2 t = \frac{1}{2} A_1 A_2 [\cos(\omega_1+\omega_2)t + \cos(\omega_1-\omega_2)t]$$

El espectro del producto, representado en la figura 16-2, **no** contiene las frecuencias originales, sino las suma y diferencia de las originales, con amplitud mitad del producto de amplitudes.

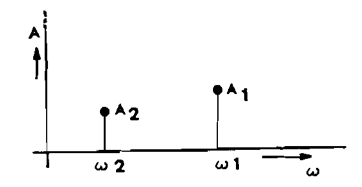


Fig. 16-1.

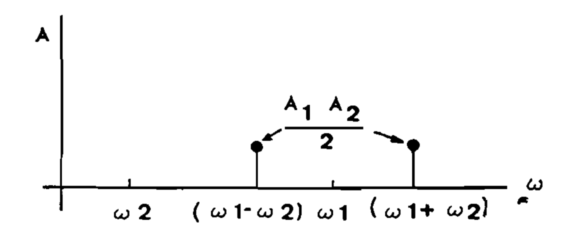


Fig. 16-2.

Supondremos $\omega_2 < \omega_1$ y contemplaremos la figura 16-2 como la "aparición de dos frecuencias por encima y por debajo de ω_1 y a la distancia ω_1 " con amplitud mitad del producto y desapareciendo las frecuencias originales. A esta operación la llamaremos **desdoblamiento de ω_2 sobre ω_1**, diremos que la frecuencia menor "se ha desdoblado" sobre la mayor.

Con esta operación (multiplicación de dos señales), que, como veremos, equivale a un tipo especial de modulación, pueden crearse frecuencias **nuevas**, suma o diferencia de dos frecuencias, cosa imposible de obtener en operaciones lineales. Con un filtro posterior puede seleccionarse la suma o la diferencia si se desea.

Como veremos enseguida, esta operación corresponde a un tipo especial de modulación (de portadora suprimida). La frecuencia ω_1 es la información que modula a la portadora ω_2. Las nuevas

frecuencias suma y diferencia se llaman **bandas laterales.**

16.3.- Producto de una onda cualquiera por otra cosinusoidal.

Sea f(t) una función del tiempo cualquiera cuyo espectro es $F(\omega)$, es decir que:

$$f(t) = \Sigma\ F(n\omega)\ \cos\ (\ n\omega t + \varphi_n\ )$$

donde $F(\omega)$ y $\varphi(\omega)$ son el espectro en amplitud y fase, con valores en $n\omega$. Si f(t) es periódica, el espectro es discreto. Estudiaremos sólo este caso, aunque es fácil generalizar el resultado a funciones no periódicas por medio de la transformación de Fourier.

Si multiplicamos f(t) por una "portadora" A cos Ωt, obtenemos:

$$Af(t)\cos\Omega t = A\Sigma F(n\omega)\cos\Omega t\cos(n\omega t+\varphi_n) =$$
$$=\Sigma\frac{A}{2}\left[F(n\omega)\cos(\Omega t+n\omega t+\varphi_n)+F(n\omega)\cos(\Omega t-n\omega t-\varphi_n)\right]$$

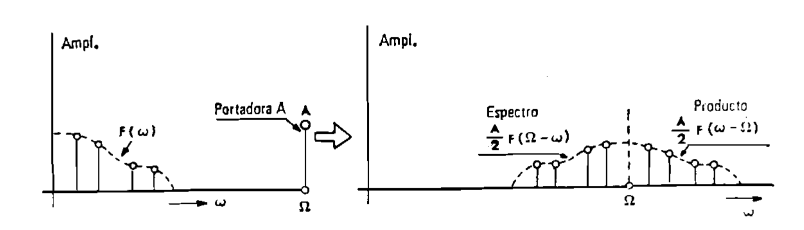


Fig. 16-3.

El espectro se ha multiplicado por A/2 y desdoblado (ha aparecido el simétrico) sobre Ω (figura 16-3). Han desaparecido las "frecuencias" Ω y las de $F(\ \omega)$ (las bajas, en este caso).

16.4.- Modulación en amplitud (AM). Bandas laterales.

Se dice que una onda **portadora** de pulsación ω_p está modulada en amplitud por la función **moduladora** f(t) cuando su valor instantáneo es :

$$e(t) = A\ [1 + m\ f(t)]\ \cos\ \omega_p t$$

donde debe cumplirse que f(t) sea menor que 1 en valor absoluto, es decir, f(t) debe estar "normalizada" a amplitud máxima 1. El factor real positivo m tal que $0 \leq m \leq 1$ se llama **factor o profundidad de modulación.**

La interpretación física de esta expresión puede verse fácilmente en la figura 16-4. Se ha supuesto m = 1. La onda

representada a la derecha es la portadora cuya amplitud varía con la ley f(t), que pasa a ser la **envolvente** de la onda modulada.

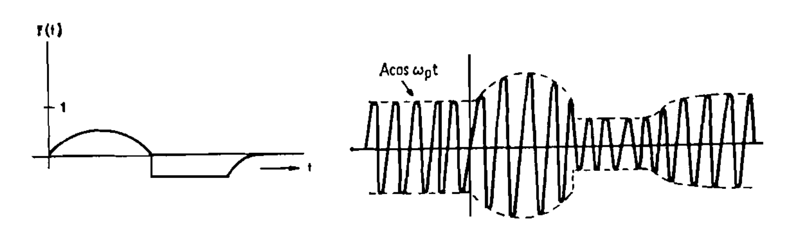


Fig. 16-4.

La nueva onda, de frecuencia próxima a ω contiene la información f(t) en la forma de variar su amplitud. Este tipo de codificación se denomina **AM** (amplitude modulation), modulación en amplitud y es el modo clásico usado en comunicaciones. El producto m f(t) debe cumplir la condición $-1 \leq$ m f(t) para que no se haga negativa la amplitud de la onda modulada, lo que implicaría un cambio de fase no deseado. Normalmente se "normaliza" que el valor máximo de f(t) sea 1 (en ambos sentidos, es decir, -1 y +1) y se elige m próximo a 1. El valor de m suele darse en %, es decir, se elige una modulación próxima al 100%.

Veamos ahora el **espectro** de la nueva onda. Para ello descomponemos previamente la función e(t) en sus dos sumandos, a saber:

$$e(t) = A\ [1 + m\ f(t)]\ \cos \omega_p t = A \cos \omega_p t + Amf(t) \cos \omega_p t$$

El primer sumando es la portadora sin modular y el segundo es un producto como el estudiado en el párrafo anterior : El espectro de f(t) se desdobla a ambos lados de la portadora. El espectro total se compone pues de la **portadora** y de dos conjuntos simétricos de **bandas laterales** que aparecen a ambos lados de la portadora, procedentes del espectro de f(t) divididos por 2). Si, por ejemplo, f(t) tiene el espectro de la figura 16-3, el espectro de la onda modulada es el de la figura 16-5.

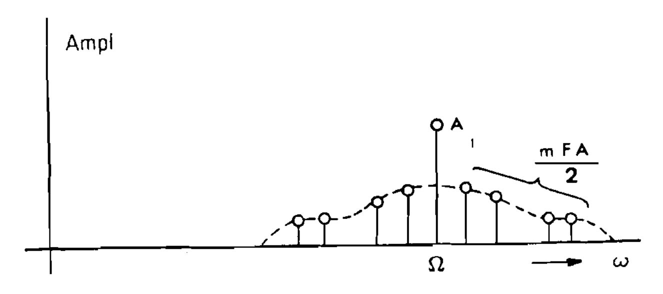


Fig. 16-5.

Veamos el caso particular de que la moduladora sea la onda sinusoidal f(t) = B cos ω_mt, que podría representar un sonido de intensidad B, de valor B<1, y de tono o frecuencia ω_m / 2π.

La onda modulada será $e(t) = A\,(1+mB\cos\omega_m t)\cos\omega_p t$, que se descompone en:

$$e(t) = A\cos\omega_p t + \frac{mAB}{2}\cos(\omega_p+\omega_m)\,t + \frac{mAB}{2}\cos(\omega_p-\omega_m)\,t$$

El espectro se compone de :

a) La **portadora**, $A\cos\omega t$.

b) Una **banda lateral superior**, de amplitud mAB/2 y a una distancia de la portadora (por encima) igual a la moduladora.

c) Una **banda lateral inferior**, de amplitud mAB/2 y a una distancia de la portadora (por debajo) igual a la moduladora.

Las figuras 16-6 y 16-7 representan la onda modulada y su espectro, respectivamente. En la figura 16-7 se han representado frecuencias en abscisas, iguales a las pulsaciones divididas por 2 π.

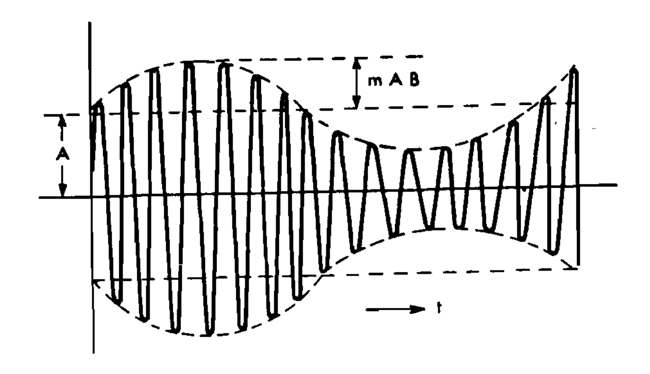


Fig. 16-6.

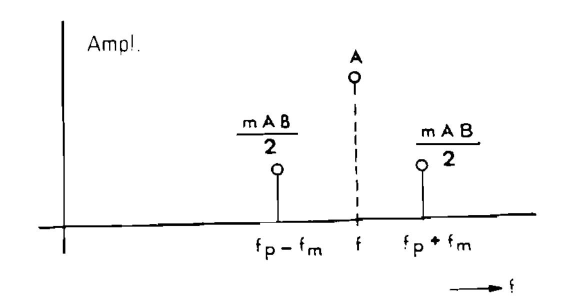


Fig. 16-7.

En las comunicaciones por modulación en amplitud, la emisora

se identifica por la frecuencia de la portadora y el programa está contenido en las bandas laterales. Si la modulación es un programa de audiofrecuencia se conviene en tomar mB próximo a 1 (por ejemplo, 0,9) para la intensidad sonora máxima. La frecuencia moduladora es igual al tono de la onda acústica. El **espectro** de la onda modulada contiene la portadora y pares de bandas laterales situados a distancias de la portadora igual a la audiofrecuencia. Generalmente se limita la audiofrecuencia a un valor máximo (por ejemplo, 5 KHz.). Las bandas laterales están siempre dentro del rectángulo dibujado en la figura 16-8. Los tonos graves poco intensos estarán situados en puntos centrales y bajos. Los tonos graves intensos en puntos centrales altos. Los tonos agudos poco intensos estarán situados en puntos bajos alejados del centro y los agudos intensos en puntos altos alejados del centro. El ancho del rectángulo, doble de la máxima frecuencia de audio es el **ancho de banda** del **canal** definido por la portadora. Normalmente, la máxima frecuencia permitida es de 5 KHz., con lo que el ancho de banda permitido es de 10 KHz. El receptor debe contener un filtro que deje pasar el ancho de banda completo y no más, para evitar que capte bandas laterales de emisoras vecinas.

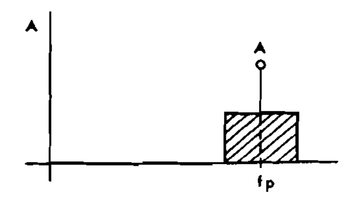


Fig. 16-8.

Las bandas laterales tienen una amplitud mAB/2. Como la potencia de una señal es proporcional al cuadrado de la amplitud, si la portadora tiene una potencia W, cada banda lateral tiene una potencia $(mB)^2$ W/4 y entre ambas $(mB)^2$ W/2, valor llamado **potencia de modulación**. Este valor tiene un máximo del 50 % de la portadora (para mB=1). **Una onda modulada al 100% contiene un 25% de la potencia de la portadora en cada banda lateral, es decir, un 50% de potencia más que la onda sin modular.**

16.5.- Obtención de la modulación en amplitud.

Para obtener una onda modulada a partir de una onda de radiofrecuencia (portadora de alta frecuencia) y otra de audiofrecuecia (moduladora de baja frecuencia) se procede de la siguiente forma (figura 16-9):

Se **suman** la portadora y la moduladora.

Se **rectifica** la suma. Con esta operación aparece una onda que puede descomponerse, como se aprecia en la figura 16-9 en la suma de la **mitad de la envolvente** y otra onda que es la **onda**

modulada, con una ligera deformación debida al rectificado.

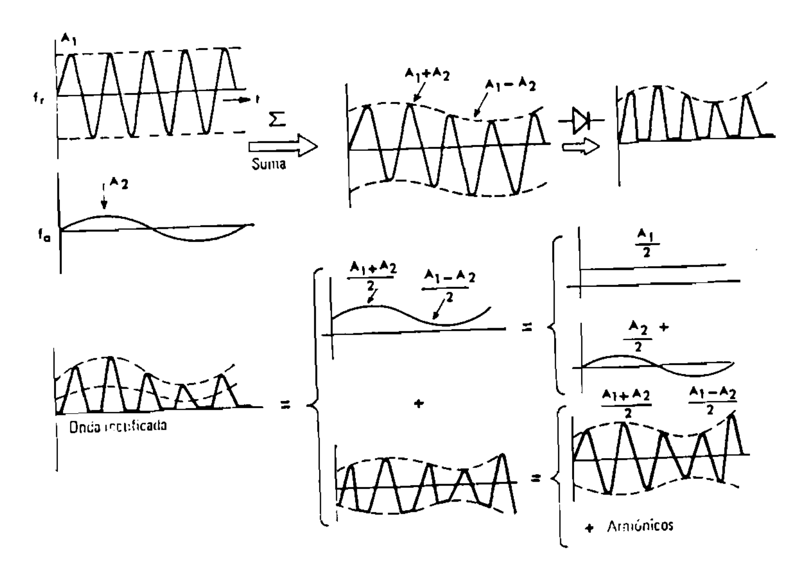


Fig. 16-9.

La onda mitad de la envolvente se descompone a su vez en una componente continua y otra de la frecuencia de la moduladora (y amplitud mitad de ésta). La onda modulada se puede descomponer a su vez en una onda modulada como la descrita en el párrafo anterior, compuesta de portadora y bandas laterales y una serie de armónicos de orden superior que son los responsables de la deformación.

El espectro de todas estas componentes es el representado en la figura 16-10 con las amplitudes indicadas en la figura, que se deducen fácilmente de la geometría de la figura 16-9 suponiendo que las amplitudes de la portadora y de la moduladora son A_1 y A_2 respectivamente.

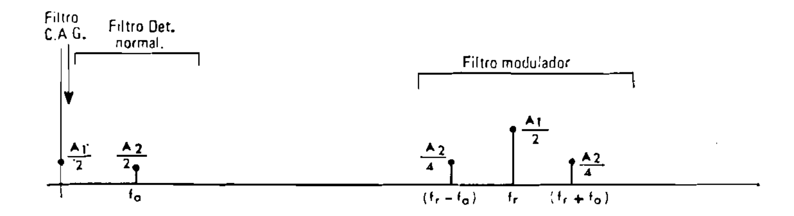


Fig. 16-10.

Con un filtro pasa-banda adecuado pueden seleccionarse la portadora y bandas laterales, es decir, la onda modulada, eliminando todas las demás. El ancho de banda del filtro debe ser el doble de la frecuencia moduladora mayor que pueda presentarse.

En comunicaciones con modulación en amplitud, se limita normativamente este ancho de banda a 10 KHz., lo que limita la frecuencia máxima del "programa" a 5 KHz.

16.6.- Detección (A.M).

Para deshacer la modulación y obtener la onda moduladora se utiliza un **detector**. El proceso de detección es el indicado en la figura 16-11.

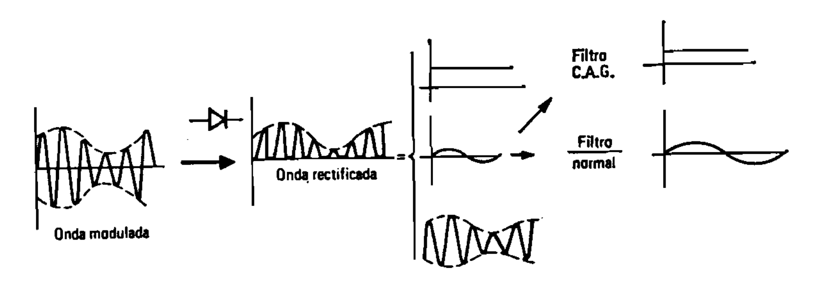


Fig. 16-11.

Primero se **rectifica** la onda modulada con un diodo. Se obtiene una onda análoga a la ya vista en la figura 16-10, es decir, un nivel de continua, la onda moduladora en valor mitad y la onda modulada, también dividida por 2, además de algunos armónicos de orden superior debidos a la deformación del rectificador.

A continuación se **filtra**, eliminando las "altas frecuencias" (las de la onda modulada) y la c.c., con lo que queda sólo la moduladora.

A veces interesa sólo la amplitud de la portadora. En ese caso basta con filtrar eliminando todas las componentes alternas y dejando sólo la c.c. Esto suele hacerse para controlar automáticamente la amplificación de los receptores según la amplitud de la señal recibida; por eso este tipo de detección se llama **c.a.g.** (control automático de ganancia).

16.7.- Conversión.

Si se desea cambiar la frecuencia de una onda f(t), modulada o no, basta con realizar el mismo proceso de modulación del párrafo 16.5, como si se fuese a modular una portadora de frecuencia f_r con otra onda f(t) de frecuencia igual al cambio de frecuencia que se desea.

El espectro de f(t) queda desdoblado y desplazado como se ve en la figura 16-10. Si ahora se filtra **seleccionando una banda lateral**, se obtiene la frecuencia suma o diferencia de las dos de entrada. Es decir, la única diferencia entre modulación y conversión es el filtro de salida. En el modulador se filtra dejando pasar las dos bandas laterales y la portadora. En el

conversor se deja pasar sólo una banda lateral, que puede ser la **suma** o la **diferencia** de las dos de entrada. El circuito de un conversor es el mismo de un modulador, siendo diferentes solamente los valores del filtro.

16.8.- Circuitos moduladores, detectores y conversores.

El modulador de A.M. debe sumar, rectificar y filtrar. La suma puede hacerse por transformador; el rectificado puede hacerse con un diodo o un transistor en clase B. En cuanto al filtro suele ser un circuito LC paralelo que actúa como filtro selectivo RLC. La figura 16-12 es un modulador con un diodo.

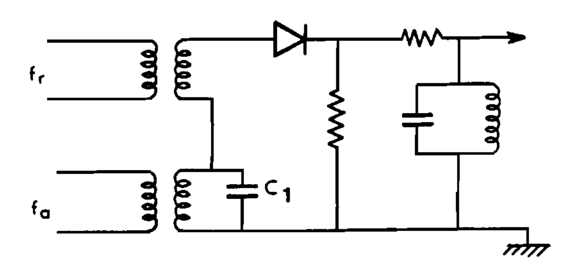


Fig. 16-12.

La figura 16-13 es un modulador en que el rectificado se hace con un amplificador de clase B, que actúa como un rectificador. Este tipo de modulación se denomina de **modulación en base**. El condensador C_1 sirve para que el secundario del transformador de audiofrecuencia presente una baja impedancia a la radiofrecuencia.

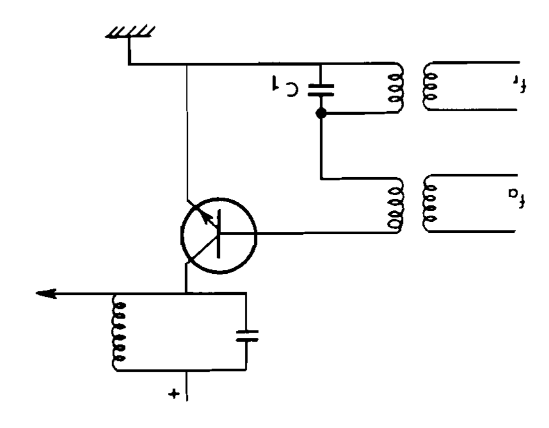


Fig. 16-13.

La figura 16-14 es un **modulador en colector**, que funciona con los mismos principios. La peculiaridad de este circuito es que suma la radiofrecuencia amplificada con la audiofrecuencia sin amplificar, lo que debe ser tenido en cuenta.

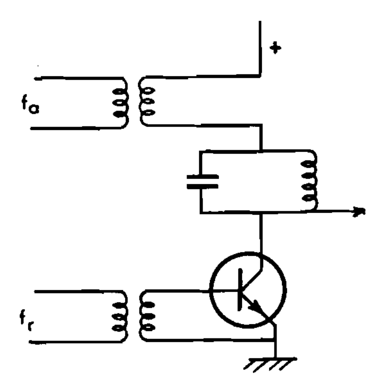


Fig. 16-14.

El **conversor** no se diferencia en nada del modulador, siendo distinto solamente el filtro de salida, como ya hemos visto.

El circuito **detector** (parte superior de la figura 16-15) debe rectificar y filtrar. El filtro "normal" (del detector de A.M.) tiene dos secciones: la primera R_1C_1 que eliminará las frecuencias superiores a la superior de audiofrecuencia y la segunda C_2R_2 que eliminará la c.c. y dejará pasar la menor audiofrecuencia presente.

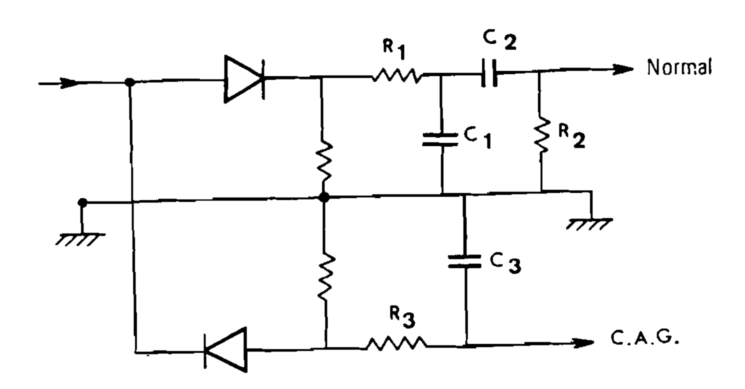


Fig. 16-15.

La parte inferior de la figura 16-15 es el **detector de C.A.G.** Su filtro debe eliminar todas las frecuencias presentes y dejar sólo la c.c.

Las condiciones que deben cumplir las componentes del detector combinado de A.M. y C.A.G., serán, si la frecuencia de audio varía entre ω_1 y ω_2 :

$$\omega_2 < \frac{1}{R_1C_1} < \omega_r - \omega_2$$
$$\frac{1}{R_2C_2} < \omega_1$$
$$\frac{1}{R_3C_3} < \omega_1$$

16.9.- Diagrama vectorial de las bandas laterales. Supresión de portadora.

A la vista del espectro de la onda modulada se aprecia que la información está contenida en las bandas laterales y, siendo éstas simétricas, en una de ellas. La onda modulada entera se podría reconstruir si se conoce una banda lateral y la frecuencia de la portadora, o, simplemente con cualquier portadora. Para comprender mejor el efecto de usar una banda lateral, veamos cómo puede representarse la onda modulada en forma vectorial:

Supongamos el caso de una onda de pulsación **Ω** modulada con profundidad **m** por una onda de pulsación ω. Sea **A** la amplitud de la portadora y **B** la de la moduladora. Representaremos la portadora por un fasor de amplitud A y fase 0, inmóvil en un plano de pulsación Ω. Esto corresponde a la imagen de un vector giratorio a Ω radianes por segundo representado sobre un papel que gira en dirección contraria a la misma velocidad angular. Dibujadas en ese mismo plano, las bandas laterales, de pulsaciones $\Omega + \omega$ y $\Omega - \omega$, aparecen como vectores giratorios a velocidades ω y $-\omega$ respecto al papel de velocidad Ω. Estos dos fasores tienen amplitud mAB/2. Son simétricos, por lo que su suma aparece siempre en fase (o contrafase) con la portadora, como se ve en la figura 16-16.

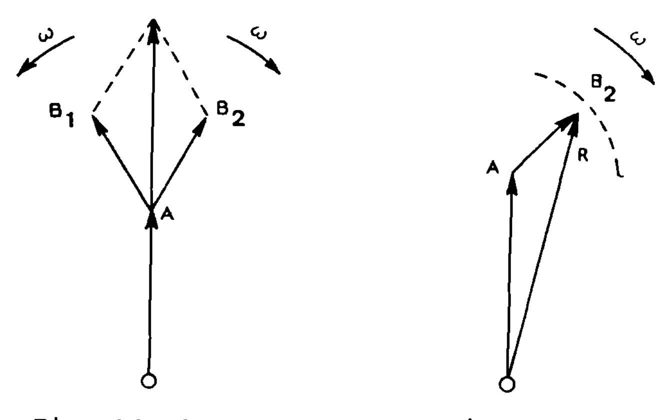


Fig. 16-16. Fig. 16-17.

Veamos ahora, a la vista de este diagrama, lo que ocurre cuando se suprime una banda lateral o la portadora:

a) **Supresión de una banda lateral.** En este caso queda el diagrama de la figura 16-17. El vector resultante R queda de amplitud variable entre A + mAB/2 y A - mAB/2, lo que equivale a que la modulación en amplitud se reduce a la mitad. Además tiene un movimiento oscilatorio que equivale a una modulación en frecuencia. La supresión de una banda lateral puede utilizarse perfectamente en comunicaciones, ya que se reduce el espectro de radiofrecuencia y se puede detectar igual que la modulación normal. Se aplica en televisión, dejando un resto de la portadora suprimida, lo que se denomina modulación "vestigial".

b) **Supresión de portadora.** En la figura 16-18 quedan sólo las dos bandas laterales. Su resultante tiene siempre la frecuencia de la portadora suprimida, si bien con medio ciclo en fase y medio en contrafase. Este efecto se aprecia mejor representando el conjunto de las bandas laterales en el plano del tiempo, como se ve en la figura 16-19.

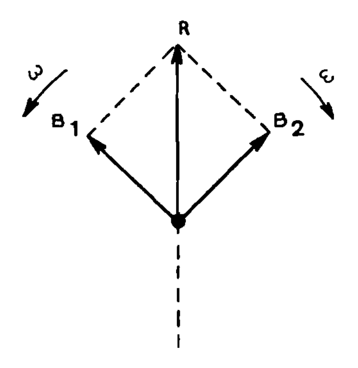


Fig. 16-18.

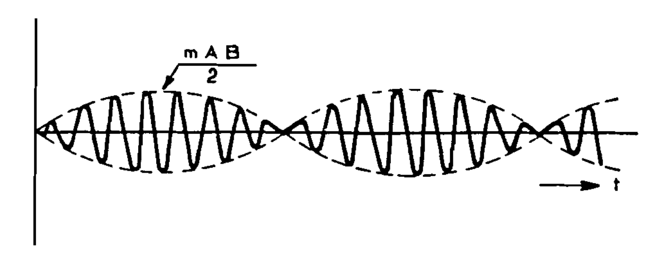


Fig. 16-19.

En la figura 16-19 se representa el conjunto de las bandas laterales. Es curioso el cambio de fase de la portadora en los ceros de la modulación. La onda "parece" modulada, aunque, en realidad, la frecuencia de la envolvente es doble de la moduladora y de forma sinusoidal rectificada.

16.10.- Aplicaciones de la supresión de B.L. o portadora.

Normalmente, las comunicaciones en A.M. se hacen con la totalidad de la onda modulada y su espectro, con portadora y dos bandas laterales.

En determinados sistemas, pueden emitirse las bandas laterales, que son las que contienen la información, con la portadora "suprimida", lo que ahorra mucha potencia en la transmisión y se reproduce la portadora en el receptor, ya que tiene un valor conocido. Este sistema se llama **DSB/SC** (double sideband/ supressed carrier).

En otros sistemas interesa usar poco espectro. Para ahorrar espectro puede suprimirse una banda lateral sin efecto aparente sobre la transmisión, como ya hemos visto. También suele suprimirse la portadora y recuperarla en el receptor, que conoce su valor y la puede "reproducir". Con ello se ahorra no sólo espectro sino también potencia de transmisión. Este sistema se llama SSB (single sideband).

En Ayudas a la Navegación se utiliza la llamada **modulación espacial**. En este sistema se transmite la portadora por una antena y las bandas laterales por otra. La modulación ocurre en el espacio. Con ello se consigue que distintos puntos del espacio tengan distinta modulación, según la proporción de señal que reciban de ambas antenas. La distribución de modulación en el espacio marca trayectorias que permiten al avión conocer su posición.

16.11.- Generador de bandas laterales.

Si se quiere generar las bandas laterales **sin portadora** se utiliza un generador de bandas laterales como el de la figura 16-20.Equivale a dos moduladores en los que la portadora se introduce con la misma fase y la moduladora con fase distinta. Al restar, se elimina la portadora. La salida debe filtrarse para eliminar la c.c. y la audiofrecuencia, como se hacía en el modulador normal.

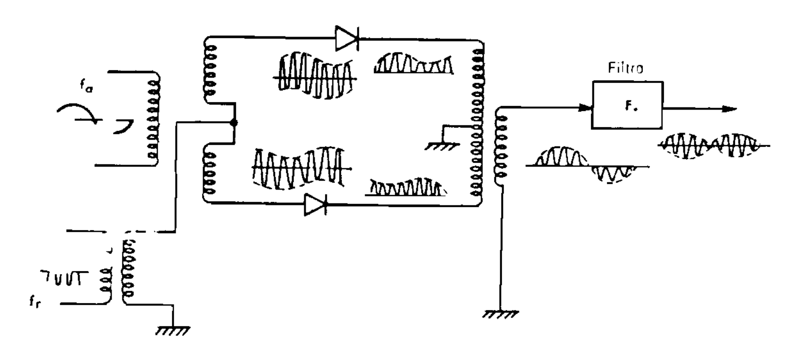


Fig. 16-20.

16.12.- Modulación en frecuencia.

La información puede codificarse en la portadora cambiando la frecuencia, en lugar de cambiar la amplitud. La onda así obtenida se llama **modulada en frecuencia**. Su expresión será:

$$e(t) = A \cos [(\Omega + mB \cos \omega t)t]$$

en que a la pulsación Ω se le suma una componente de modulación de amplitud mB (B valor de la modulación, m factor de modulación) y que varía con la pulsación de modulación ω.

Esta expresión puede escribirse de otra forma:

$$e(t) = A \cos [\Omega t + mB\, t \cos \omega t]$$

que pone de manifiesto que la modulación en frecuencia puede considerarse también como una **modulación de fase**. La fase varía sinusoidalmente multiplicada por un factor que añade una variación lineal con t, que carece de importancia.

La **desviación máxima de frecuencia** vale $mB/2\pi$ y se elige de unos 25 KHz. para el valor máximo de B.

Las comunicaciones por frecuencia modulada se hacen normalmente con portadoras de VHF o superiores, a partir de los 100 MHz.

La figura 16-21 es un circuito oscilador de reactancia variable. La frecuencia de oscilación depende de la capacidad del diodo, que es función de su polarización, que varía con la señal moduladora. Así la frecuencia de salida varía con la señal de entrada, obteniéndose una frecuencia variable.

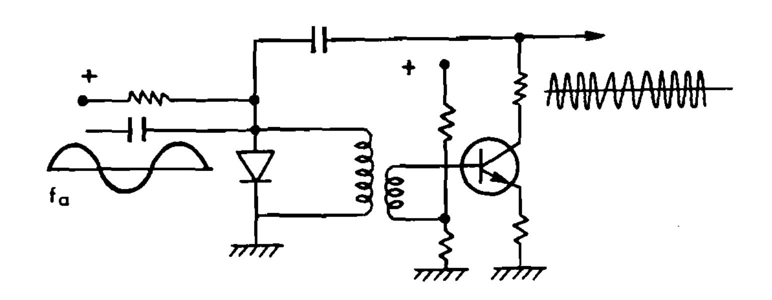


Fig. 16-21.

Para recuperar la moduladora se utiliza un **detector de FM**, también llamado **discriminador**, como el representado en la figura 16-22. Este circuito compara la frecuencia de la señal de entrada con la "nominal" o central a través de dos filtros resonantes.

Si la frecuencia "central" de la portadora es f, la de estos filtros se eligen f_1 y f_2 equidistantes de f(ver figura 16-22 y 16-23). Los dos filtros van seguidos de sendos detectores y la salida es la diferencia de niveles de señal de ambos filtros. Si la frecuencia de entrada es f, los dos filtros dan igual salida y la salida del detector, tomada entre A y B es cero.

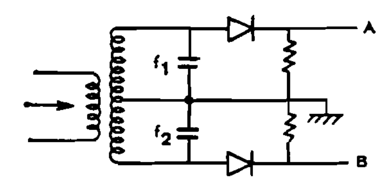


Fig. 16-22.

Si la frecuencia se desvía hacia f_1, el circuito resonante correspondiente a f_1 da más salida que el otro y la diferencia A-B es positiva. Si la frecuencia se desvía hacia f_2, el circuito correspondiente a f_2 da más salida y la diferencia A-B se hace negativa. La tensión A-B se convierte así en una medida de la desviación de frecuencia. Si la desviación de frecuencia es sinusoidal con el tiempo, A-B variará también de la misma forma. Su valor es la moduladora así recuperada.

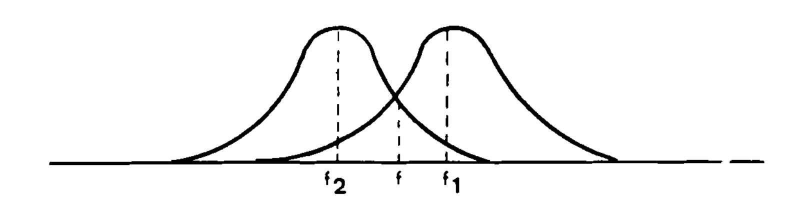


Fig. 16-23.

16.13.- Otros sistemas de modulación.

Los sistemas de modulación descritos pueden combinarse de diversas formas. A veces se utilizan **subportadoras**, que modulan a la portadora y están a su vez moduladas. Así, por ejemplo, en televisión se envía una "subportadora de sonido, de 5,5 MHz. modulada en frecuencia por el sonido. En sistemas de comunicaciones de muchos canales (llamados "de microondas" porque utilizan frecuencias portadoras muy altas) se envían hasta varios cientos de subportadoras, cada una de las cuales contiene la información de un canal, modulada en A.M. o F.M., o ambas.

La modulación de señales **digitales** con portadora analógica se hace por varios sistemas que equivalen a los anteriores, pero se caracterizan por el carácter digital de la moduladora. La moduladora es una señal de dos estados, binaria, y el modo como se diferencian esos dos estados define los diferentes sistemas de modulación digital:

En el sistema **ASK** (Amplitude shift keying) la amplitud de la portadora pasa de su valor máximo a cero, simplemente se conecta y desconecta la portadora. Equivale a una modulación AM al 100%.

En el sistema **FSK** (frquency shift keying) la frecuencia de la portadora toma un valor distinto en cada estado. Equivale a una modulación FM.

En el sistema **PSK** (phase shift keying) la fase de la portadora cambia 180º de un estado al otro. Es un caso especial, típico de la modulación digital y muy usado por su sencillez.

16.14.- Frecuencias de las comunicaciones.

La propagación de las ondas electromagnéticas es función de la frecuencia a través de las constantes eléctricas del medio, μ y ϵ. La tierra se comporta como conductora a frecuencias medias y altas, reflejándolas. En la ionosfera, la llamada "capa de Heaviside" refleja casi todas las ondas electromagnéticas que incidan con ángulos bajos. Las frecuencias medias y altas (entre 100 y 6.000 KHz.) se propagan también por "onda terrestre" que es la propagación del campo eléctrico por la discontinuidad entre la tierra y la atmósfera. Las reflexiones favorecen las comunicaciones pero son perjudiciales para las ayudas a la navegación.

Teniendo en cuenta todos estos efectos, se ha hecho una clasificación y asignación de frecuencias como se ve en la tabla de clasificación.

La propagación por "línea de visión", es decir, en línea recta, es muy fiable pero tiene el inconveniente de no llegar demasiado lejos, debido a la curvatura de la tierra.

Las reflexiones en la capa de Heaviside, típica de las comunicaciones en H.F. aumenta extraordinariamente el alcance, pero la interferencia entre ondas con diferentes trayectos reflejados produce variaciones en la amplitud de la onda recibida que varían al moverse la atmósfera, produciendo "desvanecimientos" de la señal llamados "fading". Para paliar su efecto, se recibe en varios puntos o con varias frecuencias a la vez o en "multiplex" .

Por encima de los 3.000 MHz., o sea, con **microondas**, sólo se utiliza el **radar** y las comunicaciones con satélites, para T.V. y otras aplicaciones.

FRECUENCIA	CARACTERISTICAS	APLICACIONES
De 30 a 300 KHZ. Ondas largas. Frecuencias bajas L.F.	Propagación en línea recta y por onda terrestre. Antenas muy grandes	Comunicaciones a corta distancia, AM. Ayudas a la navegación.
De 300 a 3.000 KHz. Ondas medias. M.F. Frecuencias medias.	Propagación por onda terrestre a largas distancias. Reflexiones.	Comunicaciones A.M. a media y larga distancia.
De 3.000 a 30.000 KHz. Ondas cortas. H.F. Frecuencias altas.	Propagación a larga distancia. Reflexiones.	Comunicaciones a larga distancia, A.M.
De 30 a 300 MHz. Ondas muy cortas. VHF. Frecuencias muy altas.	Propagación en línea recta. Reflexiones de objetos metálicos o conductores.	Comunicaciones a corta distancia, F.M. Televisión. Ayudas a la navegación.
De 300 a 3.000 MHz. Ondas ultracortas U.H.F.	Propagación en línea recta. Reflexiones de objetos conductores. Antenas pequeñas.	Comunicaciones a corta distancia. Televisión. Ayudas a la navegación. Radar.

16.15.- Comunicaciones A.M. y F.M.

El mensaje hablado contiene frecuencias (en su espectro) entre 300 y 3.000 Hz. La música tiene un espectro más amplio, entre 30 y 10.000 Hz, o incluso más, si se quiere transmitir con alta fidelidad. Las frecuencias entre 30 y 30.000 Hz. se llaman por eso **audiofrecuencias**. El mensaje se transforma en eléctrico a través de un micrófono, que produce una tensión eléctrica proporcional a la intensidad sonora.

Con la audiofrecuencia se modula una **portadora** de una frecuencia mucho más alta, apropiada para convertirse en onda electromagnética y enviarse al espacio a través de la antena. El receptor recibe la onda modulada a través de su antena, deshace la modulación con un detector, recuperando la audiofrecuencia que puede usarse directamente, por ejemplo, en un altavoz. Un sistema de este tipo está representado en la figura 16-24.

La modulación en amplitud, A.M. consiste, como ya hemos visto, en cambiar la amplitud de la radiofrecuencia según la amplitud de la audiofrecuencia. La onda modulada resultante se

compone de tres ondas simples, portadora y dos bandas laterales.

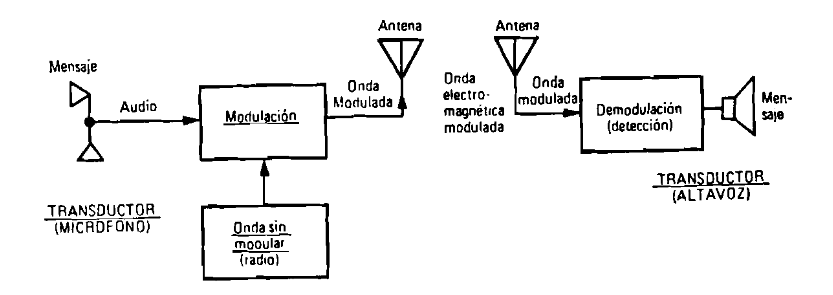


Fig. 16-24.

La modulación en frecuencia o "frecuencia modulada", F.M. consiste en variar la frecuencia según la amplitud de la moduladora. La modulación en frecuencia es menos sensible al ruido y a las perturbaciones en general que la modulación en amplitud, ya que las variaciones de amplitud no aparecen como señal.

16.16.- Transmisor A.M.

El transmisor de A.M.(fig. 16-25) se compone de:

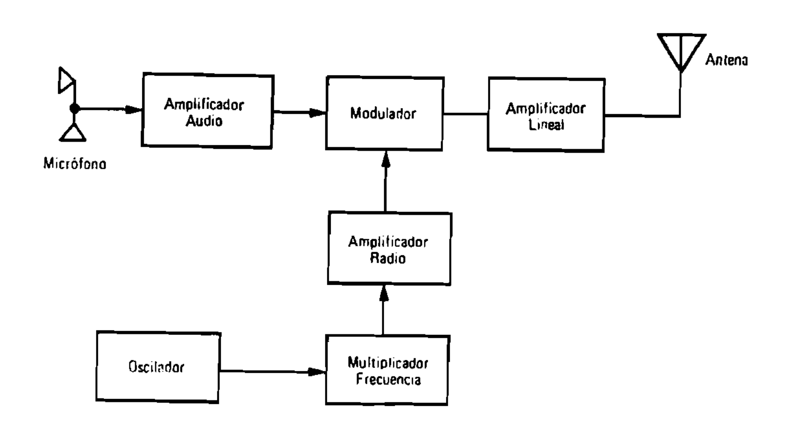


Fig. 16-25.

Un micrófono, que convierte las señales acústicas en eléctricas.

Un **amplificador de audio** en que, además de amplificar, se regula la amplitud para que la modulación sea del 90% para la máxima amplitud de señal. Esta regulación puede ser automática, reduciendo la amplificación para señales fuertes y aumentándola para las débiles, efecto que se denomina **compresor**. El amplificador debe ser lineal (salvo el efecto indicado), de

respuesta plana.

Un oscilador que genera la onda portadora. La frecuencia suele estar estabilizada con un cristal de cuarzo. A veces el oscilador se estabiliza a una frecuencia más baja que la portadora y se multiplica después la frecuencia. Para ello se distorsiona para generar armónicos y se seleccionan éstos con filtros adecuados.

Un **amplificador de radiofrecuencia** y **un modulador**, con un **amplificador** posterior, del que puede prescindirse, dependiendo de la potencia que se desea radiar.

Una antena, que transforma la onda eléctrica en electromagnética y la radia.

16.17.- Receptores de comunicaciones (A.M.).

El receptor "normal" de modulación en amplitud corresponde al esquema de la figura 16-26 y consta de:

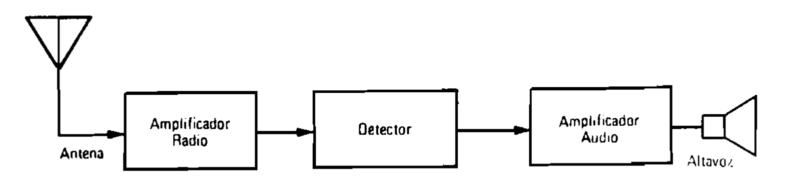


Fig. 16-26.

Una antena, que recibe la onda electromagnética y la transforma en eléctrica.

Un amplificador selectivo de radiofrecuencia, que amplifica la onda recibida y la selecciona de las indeseadas. En esta etapa se hace la **sintonía**, o elección de la emisora deseada.

Un detector, que separa la modulación dando una salida de audiofrecuencia.

Un amplificador de audiofrecuencia que obtiene la potencia de señal necesaria, y

Un altavoz, que reproduce el mensaje recibido.

Este tipo de receptor no se usa prácticamente, debido a la dificultad de sintonizar varias etapas de radiofrecuencia simultáneamente. Tampoco se usan ya otros tipos de receptor que citaremos por su valor histórico:

El receptor **reflex** utilizaba los amplificadores dos veces, haciendo pasar la audiofrecuencia por el amplificador de radiofrecuencia (apto también para frecuencias bajas). Con esta técnica se ahorraban válvulas amplificadoras.

El receptor **regenerativo** que utilizaba una realimentación positiva para aumentar la amplificación, con el consiguiente peligro de inestabilidad.

16.18.- El receptor superheterodino.

Con el esquema de la figura 16-25, la sintonía (elección de emisora que se desea recibir) se hace en etapas selectivas del amplificador de radio. Como, para que el amplificador sea eficaz, es necesario usar varias etapas selectivas, para sintonizar una emisora hay que cambiar la frecuencia de varios circuitos simultáneamente, lo que resulta prácticamente imposible.

Casi la totalidad de los receptores actuales evitan este inconveniente utilizando el principio de "mezcla" o "heterodino", por lo que se llaman **superheterodinos**. Como veremos, mediante este ingenioso procedimiento, la sintonía se hace **a frecuencia fija**.

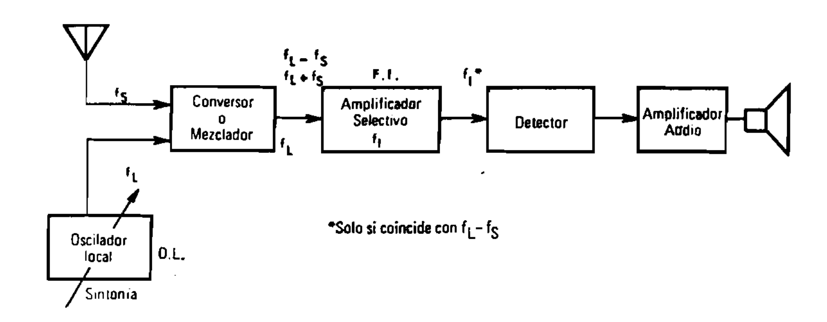


Fig. 16-27.

El receptor superheterodino consta de (fig. 16-27):

Un oscilador local sintonizable (de frecuencia ajustable con un mando). Sea f_L su frecuencia.

Una etapa **conversora** o **mezcladora** (a veces mal llamada "primer detector") en que se "bate" o "heterodina" la señal que genera el oscilador local con la que procede de la antena, obteniendo la diferencia. Se trata pues de un conversor o "mezclador" como los ya estudiados. Para cada señal que llega a la antena, de frecuencia f_S, se obtienen a la salida del conversor las frecuencias $f_L - f_S$ y $f_L + f_S$. (suponemos $f_L < f_S$).

El filtro de salida del conversor y el **amplificador de frecuencia intermedia** que sigue se sintonizan a una frecuencia f_I constante, característica del receptor, que se llama **frecuencia intermedia**. Con ello se obtiene una alta selectividad (el amplificador de F.I. puede tener varias etapas). Se **selecciona** aquella frecuencia que cumple la condición:

$$f_L - f_S = f_I$$

Las frecuencias de señal que dan diferencia con el oscilador local distinta de la frecuencia intermedia son eliminadas en el amplificador de frecuencia intermedia. Se suele elegir la frecuencia intermedia entre 400 y 500 KHz. para los receptores de frecuencias medias y de 10,5 ó 30 MHz. para los de V.H.F.). La selectividad debe estar calculada para dejar pasar la frecuencia intermedia **y las bandas laterales**.

Un **detector** separa la modulación de la frecuencia intermedia. El detector será de A.M. o F.M. según el tipo de modulación.

Por último, un **amplificador de audio** y un **altavoz** como en el caso del párrafo anterior.

El funcionamiento del receptor puede resumirse en la forma siguiente:

A la antena llegan muchas señales f_S, tantas como emisoras "presentes" con nivel suficiente. De todas ellas, si se desea seleccionar una en particular, sea f_{ST}, nosotros ajustaremos nuestra frecuencia del oscilador local f_L al valor $f_L = f_{ST}+f_I$, de forma que f_{ST} sea la única frecuencia entre las f_S que cumpla la condición $f_L - f_{ST} = f_I$. Todas las demás dan un valor distinto a f_I a la salida del conversor y son eliminadas en el amplificador de frecuencia intermedia.

16.19.- Frecuencia imagen.

La emisora seleccionada se caracteriza por cumplir la relación $f_L - f_{ST} = f_I$. Pero, desgraciadamente, por cada valor de la frecuencia seleccionada hay otro valor $f_{ST} + 2f_L$ que cumple la condición $f_S - f_L = f_I$, ya que $(f_{ST} + 2f_L) - f_L = f_I$. Esta frecuencia, simétrica de la de señal respecto al oscilador local por encima de ésta, se llama **frecuencia imagen** y, debido a que su diferencia con el oscilador local es también f_I y el conversor no discrimina el signo de la diferencia, no es eliminada en el amplificador de frecuencia intermedia. El defecto más importante del receptor superheterodino es pues que hay **dos** valores de f_S, por encima y por debajo del oscilador local, cuya diferencia con éste es la frecuencia intermedia.

Para eliminar la frecuencia imagen es necesario disponer de un **preamplificador selectivo de radiofrecuencia**, como el indicado en la figura 16-27. Si se desea sintonizar la señal de frecuencia f_{ST}, se **sintonizan simultáneamente** el amplificador de R.F. selectivo al valor f_{ST} y el oscilador local al valor $f_L = f_{ST} + f_L$.

A primera vista parece que hemos caído otra vez en el defecto de tener que sintonizar varios circuitos a la vez. Sin embargo. ese problema no es ya tan agudo debido a que la

distancia entre la frecuencia que se quiere seleccionar y su imagen es muy grande, $2f_I$, y no es necesario que el amplificador de R.F. sea muy selectivo, pudiendo tener una banda relativamente ancha. Por consiguiente, la precisión de este ajuste no necesita ser mucha. El amplificador de R.F. elimina las señales de frecuencias alejadas del orden de f_I de la que queremos seleccionar, y el amplificador de frecuencia intermedia elimina las frecuencias próximas, dejando pasar las bandas laterales, por lo que **el ancho de banda del amplificador de F.I. debe ser el doble del ancho de banda de la audiofrecuencia**. Normalmente, el ancho de banda es de unos 10 KHz., doble de los 5 KHz. necesarios para una fidelidad aceptable.

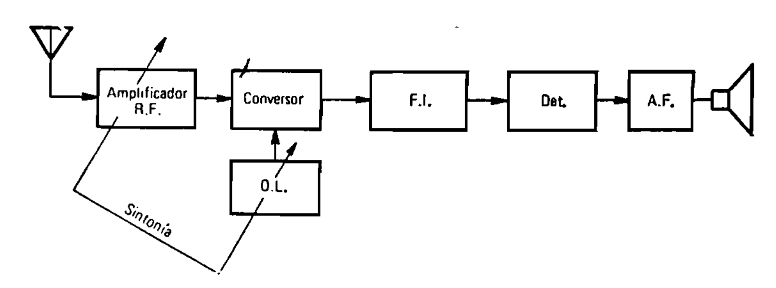


Fig. 16-28.

Todo lo explicado en este párrafo y en el anterior se refiere a receptores superheterodinos en los que la frecuencia del oscilador local se elige **por encima** de la de la señal. Sin embargo, hay algunos casos en los que la frecuencia del oscilador local se elige por debajo de la de la señal esperada, dando una frecuencia imagen más baja. No creemos necesario explicar con detalle esta posibilidad que, por otra parte, es menos usada.

16.20.- Dispositivos especiales de los receptores.

La mayoría de los receptores tienen los siguientes dispositivos:

Un **mando o control de sintonía**, que actúa variando simultáneamente sobre el amplificador de R.F. y el oscilador local, como se ve en la figura 16-28.

Un **mando o control de volumen**, que suele ser un divisor de tensión variable con un potenciómetro, a la salida del detector o en el amplificador de audio (fig. 16-29).

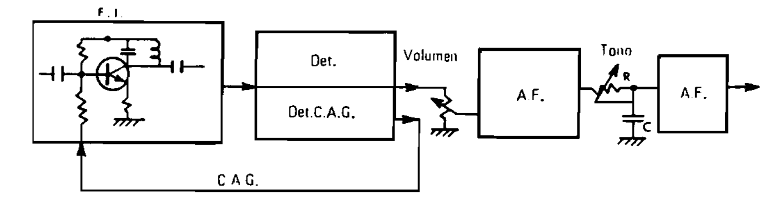


Fig. 16-29.

Un **mando o control de tono**, que cambia la constante RC de un divisor de tensión para amortiguar los tonos agudos (fig. 16-28).

Un **control automático de ganancia, C.A.G.**, que cambia la amplificación de la frecuencia intermedia en función del valor medio de la portadora. Debido a la enorme diferencia de amplitud de señal entre emisoras próximas y alejadas, resultaría muy incómoda la sintonía cuando hay muchas emisoras presentes. El control automático de ganancia toma la amplitud de la portadora del detector C.A.G. y la utiliza para disminuir la amplificación de una etapa de F.I., cambiando la polarización, como se ve en la figura 16-29. Cuando una emisora es muy fuerte, la señal del C.A.G. es muy negativa, y la amplificación de F.I. disminuye, disminuyendo también la señal en un cierto grado. Se trata de un sistema de realimentación o servomecanismo electrónico, como los que estudiaremos más adelante.

El efecto del C.A.G. se nota al sintonizar porque cuando no hay portadora presente la amplificación es máxima, y aparece un fuerte ruido de fondo. Cuando aparece una portadora, aunque sea sin programa o en un silencio del mismo, la amplificación disminuye y el ruido de fondo también. Instintivamente se selecciona mejor el receptor buscando esta ausencia de ruido que con la modulación, que es muy variable.

Todavía hay otros dispositivos especiales de los receptores como **ensanches de banda**, que cambian el ancho de banda de la F.I. para obtener más fidelidad (a cambio de perder selectividad), **filtros de audio**, que cambian la respuesta de frecuencia de audio para obtener efectos acústicos diversos, etc...

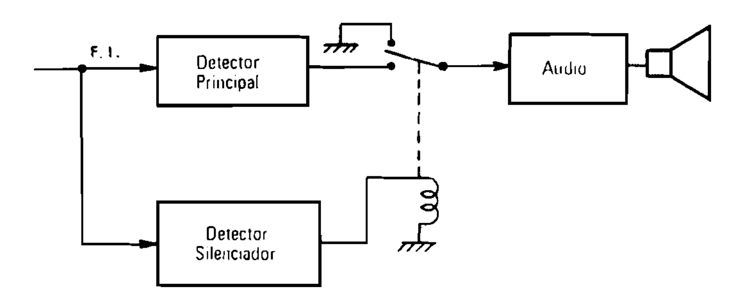


Fig. 16-30.

En las comunicaciones aeronáuticas se usa mucho el **silenciador**, que es un dispositivo que conecta el altavoz solamente cuando hay portadora, y lo desconecta el resto del tiempo, evitando así la molestia de oír todo el tiempo el ruido de fondo (fig. 16-30).

16.21.- Sistemas de modulación en frecuencia (F.M.).

El sistema de frecuencia modulada difiere del anteriormente descrito en el modulador del transmisor y en el detector del

receptor, siendo igual el resto del sistema.

El modulador consiste en un oscilador con una **reactancia variable**, como los ya estudiados, cuyo valor depende de la audiofrecuencia aplicada. Generalmente se trata de un semiconductor cuya capacidad depende de la corriente o de la tensión aplicada y que forma parte del circuito LC.

En los transmisores modernos se sintetiza digitalmente la frecuencia emisora, haciéndola función de la audiofrecuencia.

El detector del receptor se sustituye por un **discriminador**, ya estudiado, y cuyo circuito reproducimos en la figura 16-31 por comodidad.

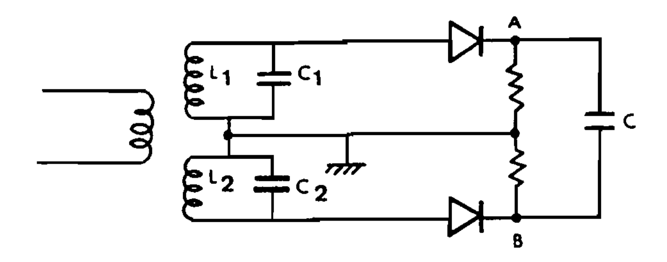


Fig. 16-31.

El circuito L_1 C_1 se sintoniza a una frecuencia superior a la central (o sea a la F.I.), como se ve en la curva A de la figura 16-32, que también representa la tensión rectificada en el punto A de la figura anterior. El circuito L_2 C_2 se sintoniza por debajo de la frecuencia "central", curva B que también representa la tensión continua en el punto B.

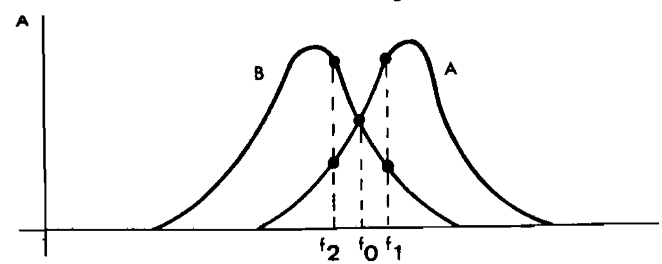


Fig. 16-32.

La carga del condensador C es la diferencia entre las tensiones A Y B, que es cero para la señal de frecuencia central f_0, positiva para frecuencias mayores (por ejemplo, f_1) y negativa para frecuencias menores que la central. La curva de salida (carga de C) del discriminador en función de la frecuencia está representada en la figura 16-33.

La salida A-B puede llevarse a un amplificador diferencial, que amplificará esta tensión que, por ser proporcional a la

desviación de frecuencia, es la audiofrecuencia. La dependencia de A-B de la frecuencia es aproximadamente lineal, si se usa sólo el discriminador para pequeñas desviaciones.

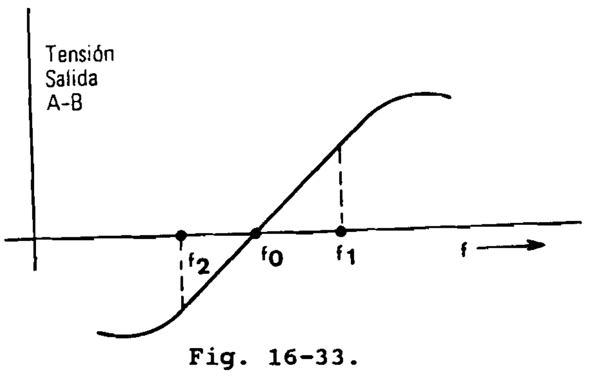


Fig. 16-33.

En los circuitos reales las dos resonancias se agrupan a veces en forma más o menos ingeniosa con el primario y secundario de un transformador, de forma que en el circuito no se ven separados como en la figura 16-31, pero el principio es el mismo.

CAPITULO 17

PRESENTACIONES EN PANTALLA

17.1.- Introducción.

La información elaborada se presenta al operador mediante **presentación de datos**. Es obvio que dicha presentación puede hacerse con altavoces, impresoras, instrumentos de indicación visual, etc...

El procedimiento visual más usado en radar, televisión, osciloscopios, ordenadores, etc... es el tubo de rayos catódicos o **pantalla**. La pantalla acumula gran cantidad de información en una superficie relativamente pequeña y, además, la presenta de forma muy elaborada, fácil de interpretar.

17.2.- El tubo de rayos catódicos.

Un tubo de rayos catódicos consta de (figura 17-1):

Un **cátodo** calentado por un **filamento** que emite un haz de electrones.

Una **rejilla de control**, con polarización negativa, que controla el **brillo** final.

Un primer **ánodo o acelerador**, de forma cilíndrica, con polarización positiva que actúa como la placa de una válvula triodo, acelerando los electrones que convierten en energía cinética y, por lo tanto, velocidad, la energía procedente de la diferencia de potencial entre el cátodo el ánodo. Sin embargo, y a diferencia con las válvulas clásicas, el ánodo es solo acelerador y no recibe los electrones, que continúan su camino hacia la pantalla.

Una **bobina de enfoque**, que cambia la dirección de los electrones enfocando el haz divergente (cambia la componente radial de la velocidad, haciéndola girar hacia dentro).

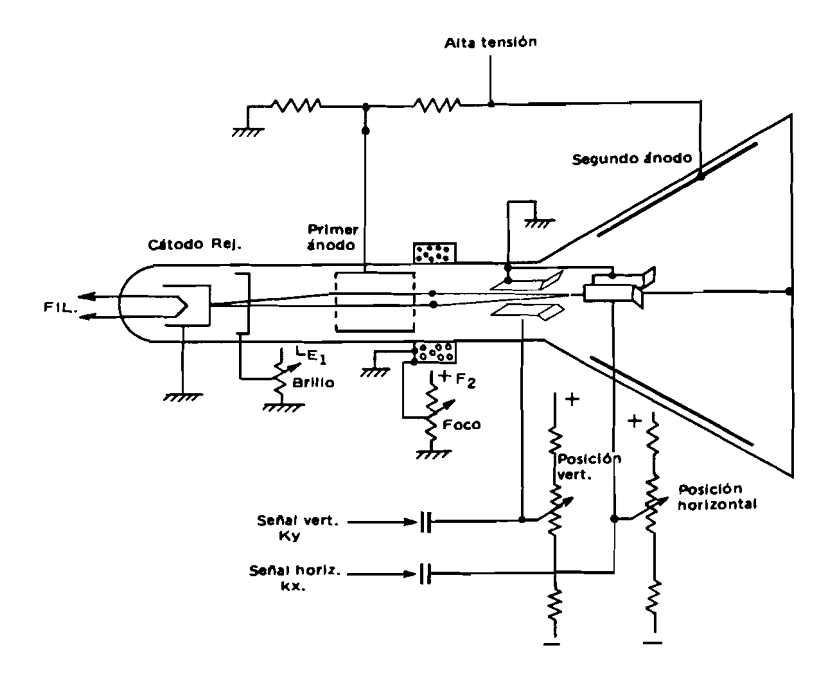


Fig. 17-1.

Unas **placas de deflexión (tubo electrostático)** que, mediante un campo eléctrico lateral, desvían el haz en un sistema ortogonal de coordenadas. La deflexión puede hacerse mediante un campo magnético perpendicular que desvía los electrones teniéndose en ese caso un tubo **de deflexión magnética.**

Un **segundo ánodo** polarizado con alta tensión que acelera más los electrones.

Una **pantalla fluorescente que se hace luminosa en el punto donde recibe los electrones.**

La tensión del ánodo se obtiene de una fuente de alta tensión. El acelerador o primer ánodo se polariza a una fracción de esta tensión.

Los controles del tubo son los siguientes (ver fig. 17-2):

La rejilla recibe una polarización variable que corresponde al control de brillo. El enfoque se regula variando la corriente de la bobina de enfoque.

Los circuitos de desviación comprenden generalmente la fijación de un valor centrado con un mando de posición y una desviación adicional, que en la figura está representada para señales alternas, pero que podría ser c.c. añadida.

Las **entradas** del tubo de rayos catódicos son **tres** tensiones, x,y,z.(figura 17-2) que se aplican a las dos desviaciones, horizontal y vertical, y al brillo. La desviación horizontal será K_x x, siendo K_x la constante de proporcionalidad (centímetros por voltio, por ejemplo) o **escala horizontal**. Del mismo modo, la desviación vertical será K_y y (con **escala vertical** K_y). El brillo será proporcional a la **entrada de brillo**, que suele ser simplemente digital, es decir, se ilumina cuando interesa que aparezca la imagen y se "borra" el trazo en el retorno.

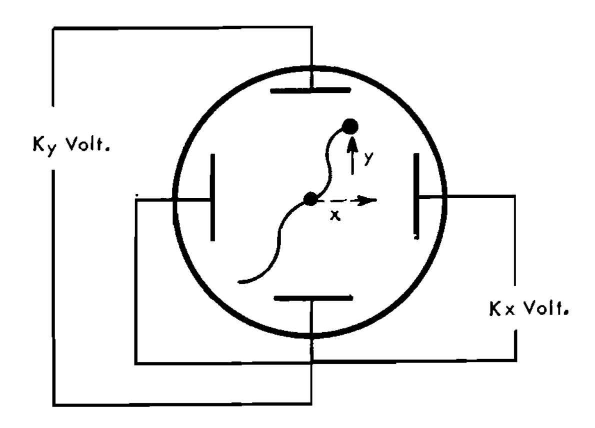


Fig. 17-2.

Cuando los valores x e y corresponden a una curva en paramétricas (x(t) e y(t)), la imagen **dibuja la curva**, que queda por persistencia de la luminiscencia si el dibujo es suficientemente rápido.

17.3.- Velocidad de los electrones y desviación.

Si suponemos el tubo simplificado a un ánodo (figura 17-3) de potencial E_1 y unas placas de desviación a diferencia de potencial E_2, tendremos:

El electrón de carga q efectúa un trabajo qE_1 al recorrer el espacio l_1. Esta energía se transformará en energía cinética, de donde se deduce fácilmente la velocidad adquirida. Si llamamos m a la masa del electrón y V_1 la velocidad adquirida, ésta será:

$$\frac{1}{2}mV^2=qE_1 \quad o\ sea \quad V=\sqrt{\frac{2E_1q}{m}}$$

suponiendo que V está lejos de la velocidad de la luz y es despreciable la pérdida de masa con la velocidad.

A partir del ánodo, la velocidad axial del **haz** es constante. Si ahora hacemos pasar el haz entre dos placas de diferencia de

potencial E_2 distanciadas l entre sí durante un espacio l_2 (ver figura 17-3) el tiempo de exposición de los electrones a este campo lateral (de valor E_2/l) será igual al espacio dividido por la velocidad y valdrá:

$$t=l_2\sqrt{\frac{m}{2qE_1}}$$

La velocidad lateral adquirida será el producto de la aceleración (fuerza del campo dividida por la masa del electrón) por el tiempo.

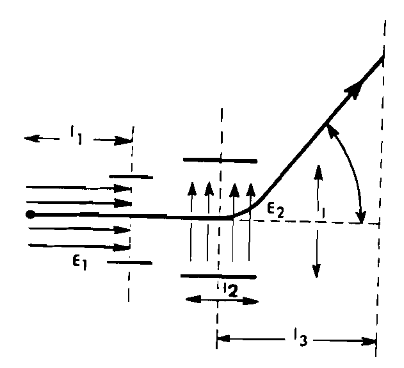


Fig.17-3.

A partir del valor de la velocidad lateral, se calcula fácilmente el ángulo de desviación α (ver la fig. 17-3), dividiéndola por la velocidad axial. Es decir:

$$V_{lat}=\frac{E_2 q}{lm}l_2\sqrt{\frac{m}{2qE_1}}=\frac{l_2 E_2\sqrt{q}}{l\sqrt{2mE_1}}$$

$$tg\alpha=\frac{V_{lat}}{V}=\frac{l_2 E_2}{2lE_1}$$

Los electrones describen una parábola entre los electrodos de desviación y salen como si procediesen del centro (recuérdese la propiedad geométrica de la parábola de que la tangente corta al eje horizontal a la mitad de la abscisa). La desviación y después de recorrer el espacio l_3 será:

$$y=\frac{l_3 l_2}{2l}\frac{E_2}{E_1}=KE_2$$

proporcional a E_2.

También se puede provocar la desviación mediante un campo magnético perpendicular. Los electrones se ven sometidos a una aceleración perpendicular a V y al campo magnético y proporcional al producto. El efecto es similar al anteriormente calculado.

17.4.- El osciloscopio.

Un osciloscopio es un instrumento electrónico para la observación de tensiones eléctricas funciones del tiempo, basado en un tubo de rayos catódicos cuyas entradas x, y, z son accesibles y se puede regular con mandos exteriores el valor de las constantes de amplificación de x e y, así como programar el brillo.

El **modo** de funcionamiento **xy** consiste en hacer una representación x-y con brillo constante.

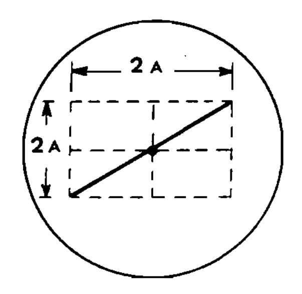
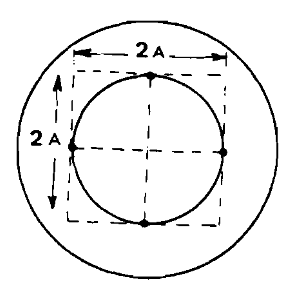


Fig. 17-4.

Veamos algunas combinaciones sencillas de x(t) e y(t). Sea x(t) = y(t) = A sen ωt. El rayo se moverá sinusoidalmente a lo largo de la recta x = y. El valor máximo será x=y=A, siempre a la escala cm./V. que corresponda a la posición de los mandos. Véase fig. 17-4.

Sea x(t) = A cos ωt, y(t) = A sen ωt. El punto x, y se mueve (fig. 17-5) sobre un círculo (que será una elipse si las escalas horizontal y vertical son distintas).

Sea ahora x(t) = A cos ωt, y(t) = A sen nωt (fig. 17-6). Al recorrer x(t) un ciclo, y(t) recorre varias veces el espacio entre A y -A. La figura se repite con la frecuencia de pulsación ω. La figura obtenida se llama de "Lissajous". En la figura se ha representado el caso n=3. Con valores distintos de frecuencias de x e y de relación racional se obtienen figuras de Lissajous de orden superior. Si la relación no es racional, las figuras aparecen móviles. Si se utiliza también la entrada z, las curvas aparecen punteadas a la frecuencia que corresponda.



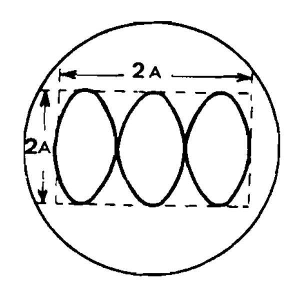


Fig. 17-5. Fig. 17-6.

Sin embargo, el **modo** más usado en el osciloscopio es el **yt**, especialmente interesante para la observación de **señales periódicas.**

Para observar una **onda periódica** y(t) se hace lo siguiente:

a) Se genera una onda en diente de sierra del mismo periodo que la que se quiere observar. Esto se hace con los **circuitos de barrido.** Esta onda será x(t), (fig. 17-7).

b) Se hace que el diente de sierra comience con el periodo. Esto se llama **sincronizar.**

La onda en diente de sierra equivale a x=Kt en su parte lineal, siendo t el tiempo **desde el principio del periodo**, y K la amplificación horizontal. O sea que las abscisas representarán el tiempo con la escala K, en microsegundos/cm., por ejemplo. El valor de y(t) en cada momento del periodo aparecerá con la escala vertical. La figura que aparecerá **repetitivamente** en la pantalla será el periodo de y(t), como se ve en la figura 17-7.

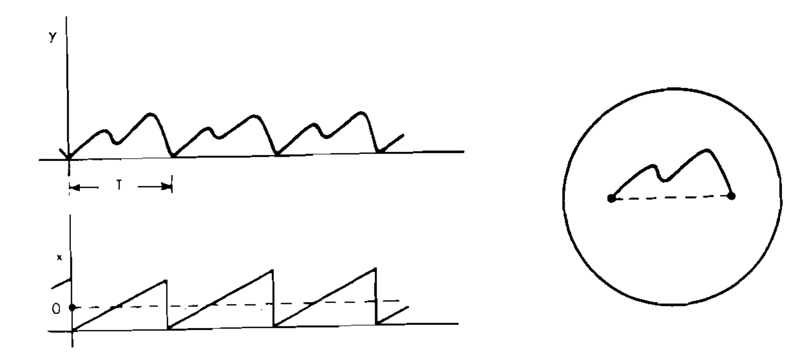


Fig. 17-7.

En general no se hace que el **barrido** tenga el mismo periodo que la que se quiere observar, sino que se dan pendientes

(escalas en microsegundos por centímetro) normalizada, de forma que la desviación máxima de x coincida aproximadamente con el periodo. Con eso se leen más fácilmente los valores de las abscisas (del tiempo). Sin embargo debe conservarse exactamente la sincronización. En la pantalla aparecerá un fragmento del periodo, como se ve en la figura 17-8. En este ejemplo se ha tomado un periodo de 8 microsegundos y un barrido de 10.

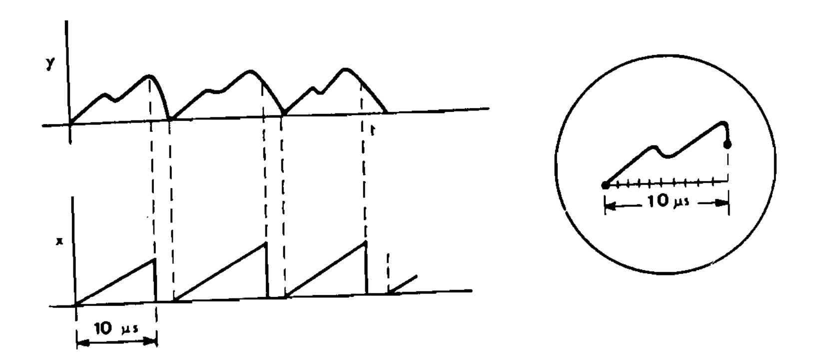


Fig. 17-8.

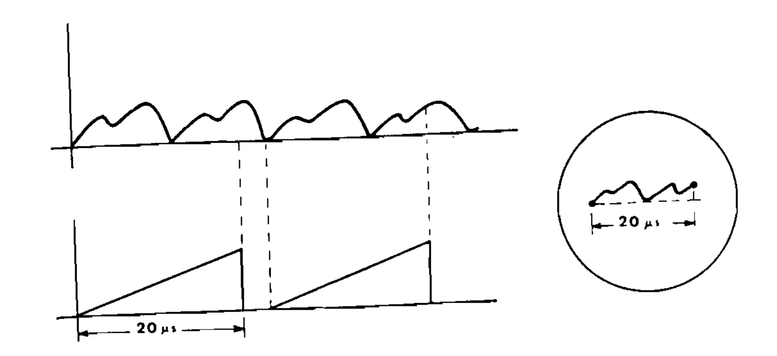


Fig. 17-9.

La figura 17-9 representa el caso de la misma onda anterior de unos 12 microsegundos con un barrido más largo de 20 microsegundos, pero también correctamente sincronizado. Si no está bien sincronizado, la onda se mueve, porque no comienza siempre en el mismo punto.

17.5.- Barrido.

El circuito de barrido corresponde al de diente de sierra, que ya hemos descrito. La carga del condensador se inicia cuando la y alcanza un valor determinado (sincronización a mano) o un valor pequeño distinto de cero (sincronización automática). La figura 17-10 es un esquema de bloques del circuito de barrido para el ejemplo de la figura 17-9. El impulso rectangular o

"puerta" generado puede utilizarse a través de la entrada z para iluminar el tubo durante el barrido y mantenerlo oscuro el resto del tiempo.

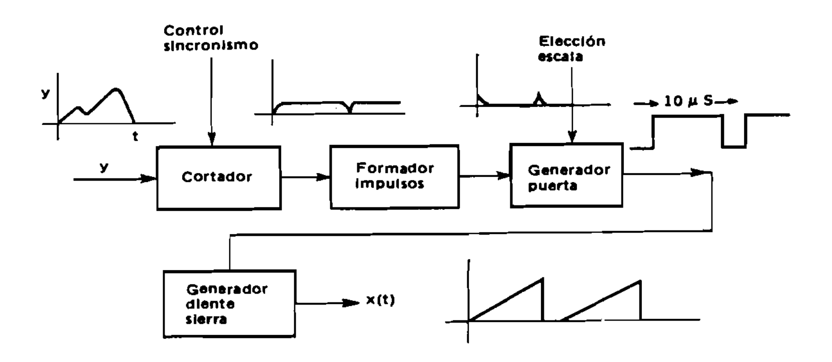


Fig. 17-10.

Si el tubo es de barrido magnético, hay que contar con la autoinducción L de la bobina de desviación. Para obtener un diente de sierra de intensidad a través de una bobina de autoinducción L y resistencia R, es necesario dar una tensión IR + LdI/dt que tiene la forma que se ve en la figura 17-11. Esta forma se obtiene por superposición de un impulso cuadrado o de "puerta" y uno de diente de sierra.

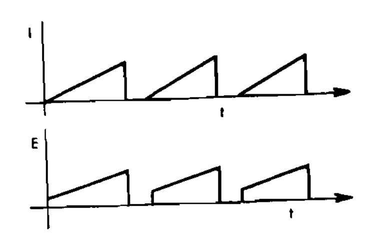


Fig. 17-11.

17.6.- La imagen de televisión (norma europea).

Para transmitir la imagen de televisión se hace recorrer al rayo luminoso la pantalla mediante un barrido en zig-zag que la cubre por completo por líneas sucesivas (fig. 17-12). La norma europea es de 625 líneas que se recorren en dos **semicuadros** de 312,5 líneas. Primero se recorren líneas alternadas comenzando en el centro superior (líneas rayadas de la figura) y luego se recorren las restantes (en blanco en la figura) empezando en la esquina superior izquierda y terminando en el centro inferior. Se transmiten 25 imágenes por segundo, es decir, 50 semicuadros de 312,5 líneas. Con este **entrelazado** se pretende disminuir la sensación de "parpadeo" que darían 25 cuadros por segundo.

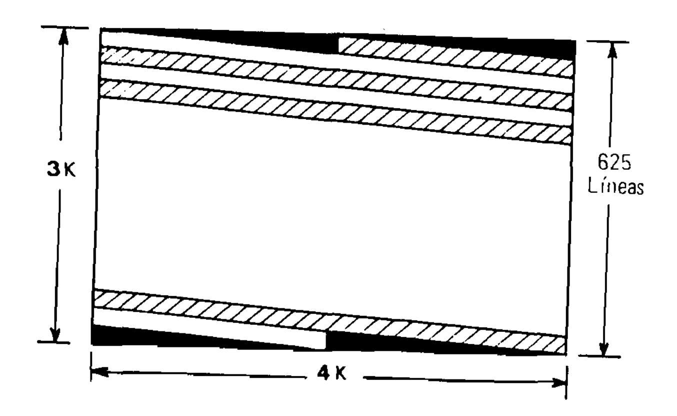


Fig. 17-12.

El rectángulo tiene un alargamiento de 4/3, que supone que si lo dividimos en puntos cuadrados, la imagen de 625 líneas tendría 625 x 625 x 4/3 = 520.833 puntos. Si se recorren estos puntos 25 veces por segundo, se debe disponer de información luminosa de 13.020.825 puntos por segundo. En el caso más desfavorable de alternar puntos blancos y negros, la señal luminosa llamada **video** sería una onda rectangular de 6,5 MHz., es decir, para mantener totalmente la resolución citada, el ancho de banda del video debería ser de 6,5 MHz. En realidad no se mantiene una calidad tan alta, es decir, la resolución o tamaño de los puntos no se hace tan pequeña. La norma europea exige un ancho de banda próximo a los 5 MHz. La información de video es la entrada z de la pantalla, que, en general suele ser un tubo de rayos catódicos de desviación magnética. (Se elige la desviación magnética porque la longitud del tubo es menor que la de los de desviación electrostática).

El barrido necesario para obtener la imagen indicada se hace con dos sistemas de dientes de sierra, llamados **horizontal** y **vertical**, que corresponden a las entradas x e y del tubo.

La desviación horizontal se hace 625 x 25 = 15.625 veces por segundo. El periodo o duración de las líneas horizontales es 15.625^{-1} = 64 microsegundos. De este tiempo se pierde el 18% en el "retorno", de forma que la línea real dura unos 52 microsegundos y el retroceso unos 12, sincronizando con **impulsos horizontales** recibidos junto con el video. El video de imagen e impulsos se recibe **invertido**, de forma que los impulsos se reciben en **negro**, lo mismo que la señal de imagen durante el retorno. La imagen sólo se ve cuando el nivel baja hacia el nivel de **blanco** o nivel mínimo, como se ve en la figura 17-3. Los impulsos de sincronismo ocupan 5,7 microsegundos, dentro de los 12 emitidos en negro.

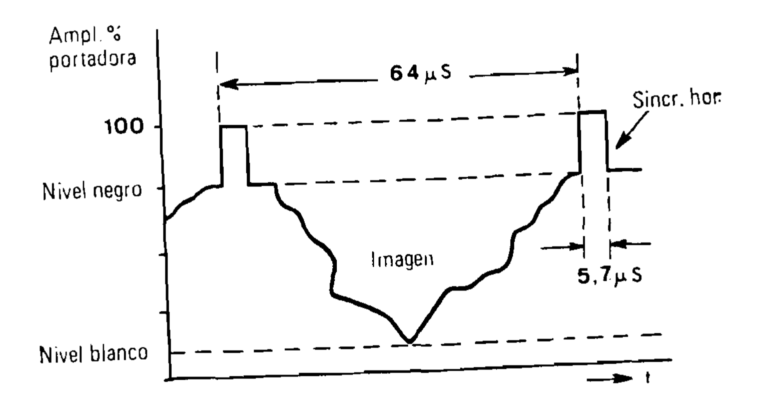


Fig. 17-13.

La desviación vertical se hace 50 veces por segundo con un periodo de 20 milisegundos. De ellos un 7% se pierde en retroceso, quedando sólo 18,6 de desviación y 1,4 de retroceso. En el retroceso (que naturalmente se hace en "negro") se pierden 40 líneas, quedando sólo 585 efectivas. Durante el retroceso vertical no se interrumpe el sincronismo horizontal, que no molesta por no aparecer en pantalla.

El sincronismo vertical se hace con 6 impulsos parecidos a los horizontales pero más anchos, de 25 microsegundos (fig.17-14). Debido a este ancho, el receptor los separa fácilmente de los horizontales, mediante un integrador, cuya descarga se facilita mediante impulsos **de igualación** más estrechos que los horizontales y que se emiten antes y después de los verticales.

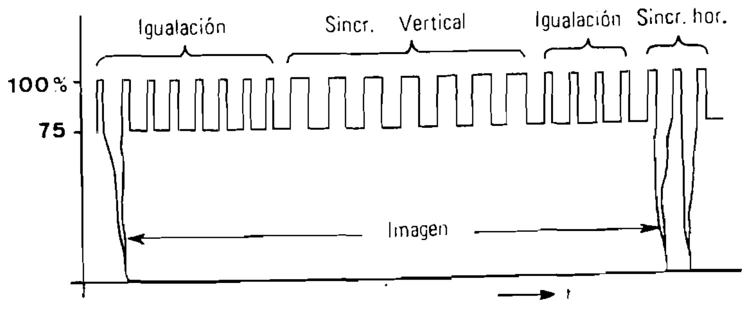


Fig. 17-14.

17.7.- La señal y el receptor de televisión.

La señal de televisión ocupa una banda en VHF y otra en UHF, localizada en canales de frecuencia establecida por las normas.

La señal de **video**, es decir, imagen e impulsos de

sincronismo se envía por modulación en amplitud y sistema **vestigial**, es decir sólo una banda lateral y un "resto" de portadora y de la otra banda, en amplitud linealmente decreciente, desde el 100% a 0,5 MHz. de la portadora y al lado correspondiente a la banda lateral única (O sea, 100% entre 5 y 0,5 desde la portadora)hasta 0 a 0,5 MHz al otro lado de la portadora, de la que queda también un 50%. El ancho de banda es de 5 MHz. La señal de video es como se indica en la figura 17-14, con un nivel de **negro** del 75% de amplitud de la portadora (los impulsos por encima en el restante 25%) y un nivel de blanco del 10%.

La información de **sonido** se envía con una **subportadora** de 5,5MHz. en A.M. y también en banda lateral única, es decir, que la "subportadora" de sonido aparece a 5,5 MHz. del mismo lado de la banda lateral de video, con lo que toda la señal cabe en el ancho de 6 MHz. asignado al canal (la distancia entre canales es de 7 MHz.). El sonido se envía modulando en F.M. esa subportadora.

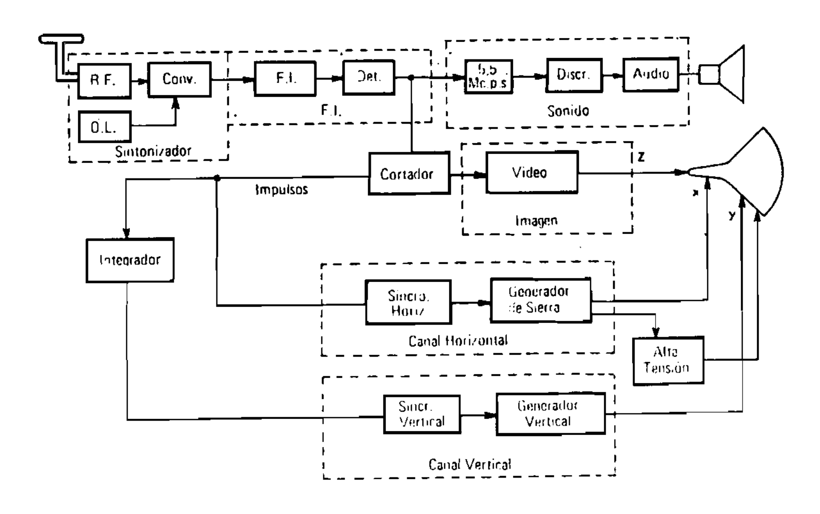


Fig. 17-15.

La figura 17-15 es un diagrama de bloques simplificado de un receptor de televisión elemental. La parte de R.F.es un receptor superheterodino. A partir del detector de A.M. se separa el canal de sonido con un filtro de 5,5 MHz. y un discriminador y el canal de video; de él se separa la imagen cortando los impulsos y los impulsos horizontales por su nivel. Los impulsos verticales se obtienen en un integrador. Tradicionalmente la alta tensión se obtiene transformando el fuerte salto de tensión de retorno de los impulsos horizontales.

Hay que señalar que aquí sólo se han descrito algunos principios básicos de televisión, para comprender mejor el uso

de los sistemas de codificación de imagen. Se han omitido detalles tan importantes como la televisión en color, la de alta resolución, el teletexto, etc..., que deben consultarse en la amplia literatura especializada en televisión.

17.8.- Otras presentaciones de datos en pantalla.

El uso de la pantalla para la presentación de datos se ha extendido a todas las ramas de la instrumentación, debido a su gran facilidad de acceso y presentación de gráficos y textos en forma parecida a la de un libro o documento.

Los modos de presentación en instrumentación se dividen en dos grandes grupos, a saber:

a) Pantalla **vectorial**. Es el sistema más directo y flexible. La imagen se compone punto a punto con los valores z(x), z(y), siendo x e y las coordenadas del punto de nivel de brillo z que se van leyendo en una memoria especial de la pantalla. El brillo se retiene entre lecturas un cierto tiempo. Si se quiere dejar la imagen fija se "refrescan" estas lecturas. La imagen puede estar escrita en cualquier orden, pero si se trata, por ejemplo, de dibujar una letra o un segmento, no se necesita "recorrer" el resto de la pantalla y se escribe como se hace con un lápiz en un papel, un punto tras otro, con más o menos brillo. Es el sistema descrito para el osciloscopio.

b) Pantalla con retícula (**raster**). Es un sistema parecido al de la televisión. El haz "barre" toda la pantalla y se ilumina donde se le ordene. También se hace con una memoria propia, pero cambia el sistema de lectura. Es más adecuado si las figuras son muy densas y complicadas, porque las variables x e y no cambian a gran velocidad, como pueden hacerlo en el sistema vectorial, pero llenan bien todo el espacio.

Por lo demás existen muchas variaciones de estos sistemas, que se han complicado con la aparición de otros tipos de pantallas que no son de rayos catódicos como las de cristal líquido, de plasma, etc...La explicación de todos estos tipos de presentaciones se sale del propósito de este libro.

CAPITULO 18

SISTEMAS DE RADAR

18.1.- Introducción.

La palabra radar procede de **radio detection and ranging** y significa la localización de objetos por medio de ondas electromagnéticas. La localización se hace orientando la antena hacia el objeto y midiendo la distancia por el tiempo que transcurre entre la salida de las ondas electromagnéticas emitidas y la vuelta del eco reflejado por el objeto, teniendo en cuenta que la velocidad de propagación es constante, igual a la velocidad de la luz.

Aunque el radar era conocido desde una patente alemana de 1904 (Hülsmeyer), no se experimentó hasta 1930 en su forma más sencilla (simple detección de la presencia de un avión) y 1935 en forma de radar de impulsos. El perfeccionamiento del radar de a bordo en 1939 fue una de las causas de la supremacía aérea británica en la segunda guerra mundial.

Una primera clasificación de los sistemas de radar es en **radar de impulsos** y **radar de onda continua** . En lo sucesivo describiremos el radar de impulsos hasta el final del capítulo en que nos ocuparemos del radar Doppler, que corresponde al de onda continua.

23.1.- Radar de impulsos.

Para medir el tiempo transcurrido entre la ida y vuelta de la onda electromagnética se emite un "impulso" de onda que quiere decir que se emite una onda sinusoidal durante un tiempo muy corto. Se puede, por ejemplo, emitir una frecuencia de 9.600 MHz. (λ = 3,1 cm., T =0,000104 μS.) durante 0'05 μS., es decir, enviar un "tren de ondas" de 480 periodos.

Al emitir el impulso se empieza a medir el tiempo para computar la distancia. Para ello se sincroniza un diente de sierra para la presentación (figura 18-1). El diente de sierra

se calcula para.que dé una desviación de, por ejemplo, 1 cm. por cada 10 μS. Con ello, y en presentación "A", si la pantalla tiene 10 cm. se desvía un total equivalente a 100 μ seg., que es el tiempo de ida y vuelta para un blanco que esté a 15 Km. Las ondas recorren 300 m. por microsegundo, es decir 1 μS. equivale a 150 m. (recuérdese que tienen que ir y volver). A este factor de **150 m. / μS.** (mitad de la velocidad de la luz) se le llama escala del radar.

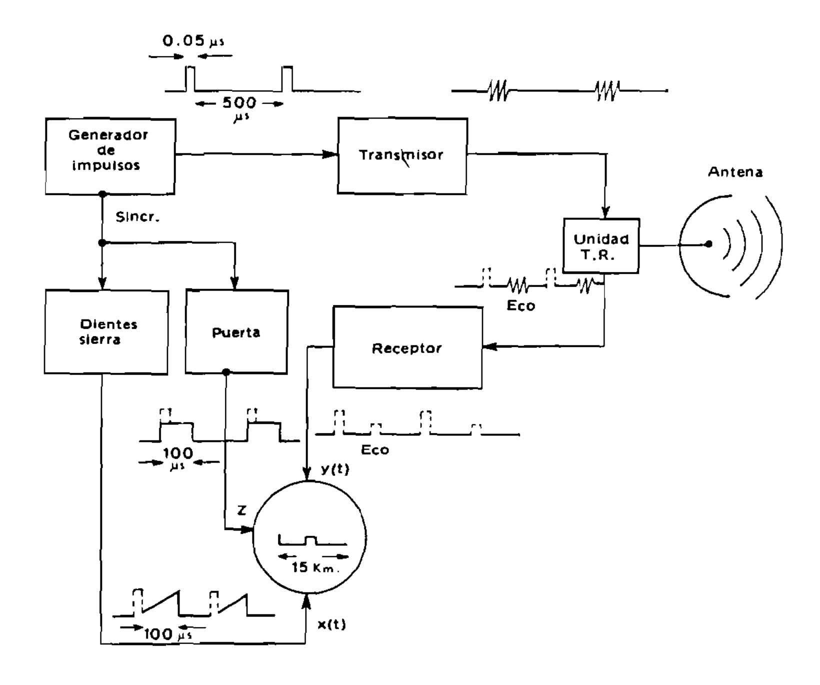


Fig. 18-1.

La desviación y de la pantalla se obtiene del receptor, corresponde a la amplitud del blanco. La desviación x se obtiene del diente de sierra, y corresponde a la distancia, a la escala del radar. El "eco" produce una desviación vertical en el momento que corresponde a su distancia, que, así, se lee fácilmente en la pantalla.

Con una sola medida no se vería nada, ya que la escritura del eco en 0'05 μS. no seria legible. La mediad se repite con un cierto ritmo de periodo superior al de llegada de los ecos más lejanos. En nuestro ejemplo puede repetirse cada 500 μS., es decir, con una **frecuencia de repetición de impulsos** (p.r.f., pulse repetition frequency) de 2.000 c.p.s. Esto garantiza la "persistencia" fluorescente de la pantalla.

La antena de transmisión y la de recepción suelen ser la misma. Para que el impulso del transmisor no pase al receptor se utiliza un conmutador electrónico llamado T-R.

Un generador de "puerta" oscurece el tubo de rayos catódicos fuera del tiempo de medida, para que no se vea la "vuelta" del haz luminoso.

23.3.- Ecuación del radar.

Si la potencia del transmisor es Pt, y si la antena fuese omnidireccional, la "densidad de potencia" a la distancia R seria:

$$\frac{P_t}{4\pi R^2}$$

Si la antena tiene una ganancia G en la dirección principal, la densidad de potencia en esa dirección estará multiplicada por G; si, además, el "blanco" tiene una superficie S, la potencia recibida por el blanco será:

$$\frac{P_t G S}{4\pi R^2}$$

Si se llama S a la superficie "proyectada" equivalente del blanco, es decir, la superficie de un blanco ideal que refleje la totalidad de la potencia en forma omnidireccional ("cross section"), a la antena receptora llegará una densidad de potencia reflejada

$$\frac{P_t G S}{(4\pi R^2)^2}$$

Si se llama A al área de la antena, la potencia recibida será:

$$P_r = \frac{P_t G S A}{(4\pi R^2)^2}$$

Si llamamos ahora N a la potencia mínima detectable por el receptor, el alcance máximo será aquél para el cual

$$N = \frac{P_t G S A}{(4\pi R^2)^2}$$

Teniendo en cuenta ahora que la antena transmisora y receptora son la misma, y que la ganancia de una antena parabólica es:

$$G=\frac{4\pi A}{\lambda^2}$$

donde λ es la longitud de onda, será:

$$N=\frac{P_t S A^2}{4\pi R^4 \lambda^2}=\frac{P_t S G^2 \lambda^2}{(4\pi)^3 R^4}$$

de donde se puede despejar el alcance máximo:

$$R=\sqrt[4]{\frac{P_t S G^2 \lambda^2}{(4\pi)^3 N}}=\sqrt[4]{\frac{P_t S A^2}{4\pi \lambda^2 N}}$$

Esta fórmula" en una u otra forma, se conoce por **ecuación del radar**, y da una idea de la dependencia de los parámetros que entran en el cómputo del alcance. Desde luego no se debe tomar como fórmula exacta porque está claro que si, por ejemplo, se cambia la longitud de onda también cambia el valor posible de Pt según las componentes que haya en el mercado y de N según la calidad de los receptores en función de la frecuencia.

En principio se puede sacar la conclusión aproximada de que el alcance varía con la raíz cuarta de la potencia, es decir, que para duplicar el alcance habría que multiplicar por 16 la potencia. Si se deja la dimensión de la antena constante, el alcance varía aumentando con la frecuencia. Esto es, quizá, la conclusión menos exacta, ya que, al aumentar la frecuencia, aumenta también N y aumenta la atenuación a través del aire húmedo (esta atenuación no aparece en la fórmula), lo que equivale a disminuir Pt. Además, a partir de cierta frecuencia se ven las gotas de lluvia y puntos de condensación, ocultando cualquier otra información.

18.4.- Frecuencias del radar.

Las frecuencias de radar se clasifican en "bandas" según la siguiente tabla (de origen militar):

UHF: 300 a 1.000 MHz.
Banda L : 1.000 a 2.000 MHz.
Banda S : 2.000 a 4.000 MHz.
Banda C : 4.000 a 8.000 MHz.
Banda X : 8.000 a 12.500 MHz.
Banda K : 12'5 a 40 GHz.
Ondas milimétricas: más de 40 GHz.

Los radares de la época de 1939 eran de la banda UHF; después se usaron frecuencias más altas, L y S. Actualmente,

todavía es la S la más normal para radares de gran alcance (desde 50 Km.). Para mayor precisión se usan frecuencias mayores, por ejemplo de la banda X. Por encima de la banda X la atenuación es muy grande y tienen menos utilización.

18.5.- Ancho de impulso, frecuencia de repetición y frecuencia.

La resolución depende del ancho de impulso, ya que la mínima distancia a la cual dos blancos aparecen como separados es la que corresponde a ese ancho; si la distancia es menor, el segundo blanco empieza a reflejar señal cuando el primero todavía lo está haciendo. Es corriente utilizar anchos de impulso de unas décimas de microsegundo para los radares de precisión y se eligen algo mayores cuando se quiere aumentar el alcance, ya que la energía devuelta por el eco es proporcional al ancho del impulso. En efecto, en el párrafo anterior habíamos hablado de una "potencia mínima detectable N" como si fuese constante; sin embargo, la potencia mínima detectable es menor si el impulso dura más tiempo (es más fácil distinguirla del ruido).

La distancia entre impulsos debe elegirse lo suficientemente grande para que no lleguen algunos ecos con retardo mayor que esa distancia, porque entonces aparecerían a una distancia falsa (ese efecto se conoce por ecos de "segundo viaje"). Por otra parte, debe ser la más corta permisible, para que el receptor reciba más potencia media procedente del eco. En general se elige del orden del doble del alcance, lo suficientemente grande para evitar el efecto de blancos de segundo viaje.

La potencia del emisor debe ser lo mayor posible, estando limitada por motivos técnicos o económicos. Actualmente es posible construir osciladores del tipo "magnetrón" de algunos miles de kilovatios. Los radares más pequeños suelen tener potencias de unos kilovatios. Claro que la potencia media es muy inferior, ya que el oscilador funciona sólo una fracción muy pequeña del tiempo. La potencia media del emisor será $Pt\ T/\ T_r$ siendo T el ancho de impulso y T_r la distancia entre impulsos.

18.6.- Osciladores y emisores de radar.

Los emisores de radar se caracterizan por su gran potencia y muy alta frecuencia. Esta última característica hace la amplificación prácticamente imposible; por ello, en radar se utiliza directamente el oscilador como generador de radiofrecuencia y desde él se alimenta la antena.

Para las frecuencias más bajas, hasta unos 1.000 MHz. todavía se utilizan amplificadores convencionales preparados como osciladores. Por encima de esa frecuencia sólo hay prácticamente dos tipos de osciladores, y de ellos sólo uno de gran potencia. Estos dos tipos son los siguientes:

a) El **Klistron reflex** es una válvula especial donde el haz de electrones que sale del cátodo y se forma a través del "acelerador" (como en el tubo de rayos catódicos) es frenado por el electrodo auxiliar llamado "repulsor" (figura 18-2); los

electrones generan unas oscilaciones que se mantienen en la cavidad resonante y refuerzan la oscilación. Se puede decir que el paso de los electrones induce un campo magnético circular que se refleja en la cavidad resonante y es devuelto en forma de potencial inducido que refuerza las variaciones de velocidad que habían generado las oscilaciones. Esta realimentación mantiene las oscilaciones del haz de electrones, que los agrupa en "paquetes", en lugar de formar un haz continuo. Las oscilaciones de la cavidad se extraen mediante una "sonda" o pequeña antena.

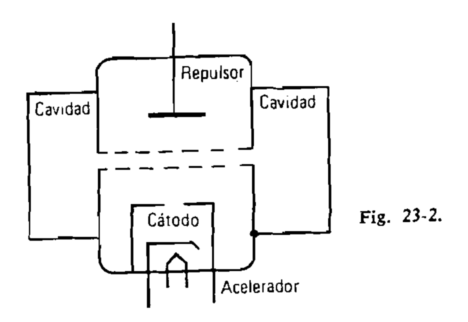


Fig. 18-2.

La figura 18-3 representa la amplitud de las oscilaciones y la frecuencia en función de la tensión del repulsor. Como se ve aparecen varios "modos" o formas de oscilar de la cavidad. La frecuencia varía ligeramente alrededor de un valor central, lo que permite un control de la frecuencia.

b) En el **magnetrón** también se hace pasar el haz de electrones por una serie de cavidades resonantes, de forma que se provoque la deseada realimentación. Los electrones, al pasar por las cavidades generan ondas (si su velocidad varía) que, reflejadas en las cavidades, refuerzan las oscilaciones. Un magnetrón consiste en un cátodo central que emite una nube de electrones

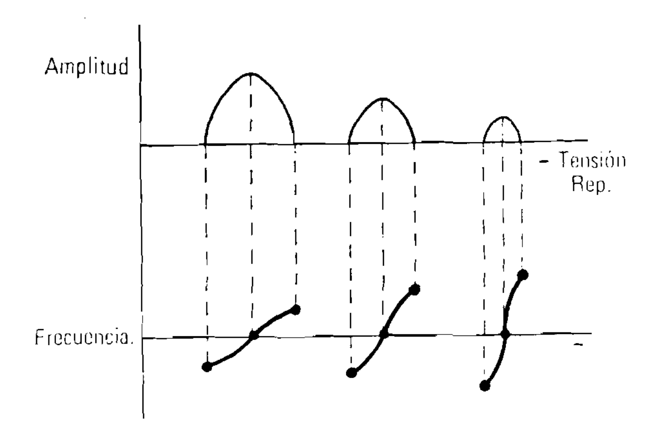


Fig. 18-3.

que son atraídos por el ánodo periférico. Un campo magnético fuerte desvía los electrones haciéndolos girar alrededor del cátodo (figura 18-4). La salida de las oscilaciones se hace con una sonda adecuada.

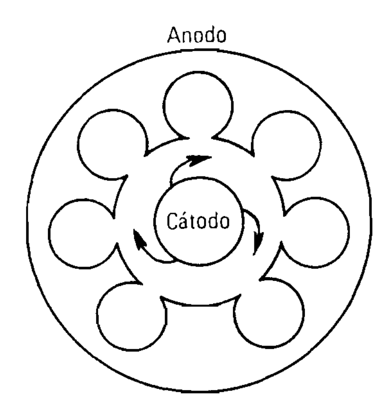


Fig. 18-4.

El magnetrón alcanza potencias muy elevadas, aunque es muy difícil de controlar o modular.

Normalmente, el emisor de radar es un magnetrón. La salida se lleva a la antena directamente con guías de onda adecuadas. El oscilador local del receptor suele ser un klistron, cuya frecuencia se ajusta para que la diferencia con el magnetrón sea la intermedia. Ultimamente han surgido algunos dispositivos de semiconductores que pueden sustituir al klistron, siempre que la potencia sea reducida. (Diodos Gunn.)

18.7.- Sistemas de radiofrecuencia.

Además de las componentes de guías de onda normales, el sistema de radar tiene dos clases de dispositivos de interés especial:

a) La unidad T-R es una cavidad resonante con una gas ionizado mediante la aplicación de alta tensión. La unidad T-R se coloca a la entrada del receptor y evita que entren las señales muy fuertes procedentes del magnetrón. Estas señales provocan una tensión muy alta entre dos puntas y hacen saltar un arco entre ellas, cortocircuitando la entrada.

b) La unión "híbrida" es una mecánica especial de guías de onda que acopla un camino y desacopla otro (ver figura 18-5). La entrada 1 se divide entre las salidas 1 y 2. La entrada 2 se divide también. Sin embargo no sale nada por cada entrada procedente de la otra, por ser los campos eléctricos perpendiculares entre sí.

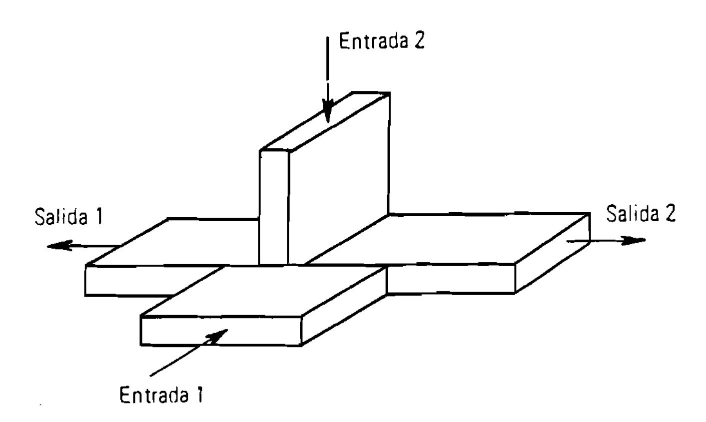


Fig. 18-5.

La figura 18-6 es un esquema de la "radiofrecuencia". La salida del magnetrón se lleva mediante una guía de onda a la antena. Las señales recibidas vienen por la misma antena; en una bifurcación de la guía de onda pasan por la "unidad T-R" (que evita que las señales del magnetrón pasen al receptor) a la unidad híbrida, donde se mezclan con las señales del klistron local. Generalmente se "sale" a través de dos detectores colocados en la misma guía de onda y cuya salida es la frecuencia intermedia (generalmente 30 o 60 MHz.). Un receptor especial mide exactamente la frecuencia intermedia y, si no es exactamente la deseada, corrige la del oscilador local. Esto se llama '~control automático de frecuencia".

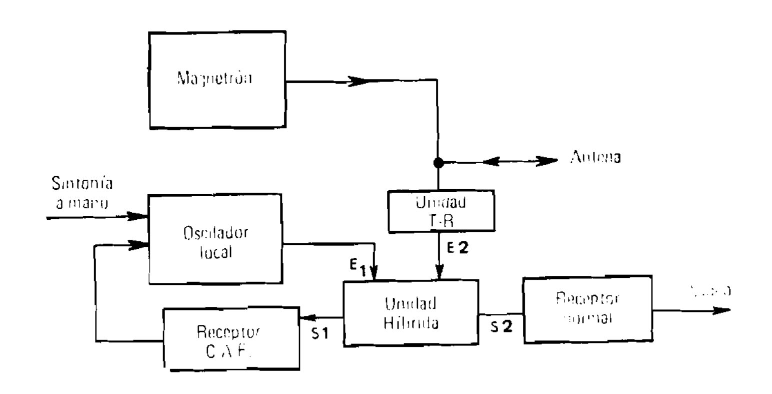


Fig. 18-6.

El receptor normal empieza ya con la frecuencia intermedia hasta extraer la modulación o "video".

18.8.- Receptores de radar.

La señal de alta frecuencia no se puede amplificar (excepto con algunos sistemas especiales de amplificadores paramétricos y maser). Por consiguiente se pasa inmediatamente a frecuencia intermedia (como en cualquier receptor superheterodino, pero sin amplificación previa) mezclando en la propia guía de onda y rectificando con un detector que es un diodo de "cristal" de silicio colocado sobre la guía (figura 18-7). A partir de ahí se trata de un receptor de bajo nivel de ruido. Sin embargo, suele tener los siguientes dispositivos especiales:

Un control de ganancia con el tiempo (figura 18-8) que disminuye la amplificación de los blancos próximos, que, naturalmente, aparecen con una amplitud mucho mayor que los lejanos, aunque sean muy pequeños.

Un diferenciador de salida, que deja sólo los bordes de los blancos muy grandes y permite ver blancos superpuestos (ver figura 18-9).

18.9.- Radar Doppler.

Si se quiere medir la velocidad del blanco, se utiliza el llamado radar Doppler.

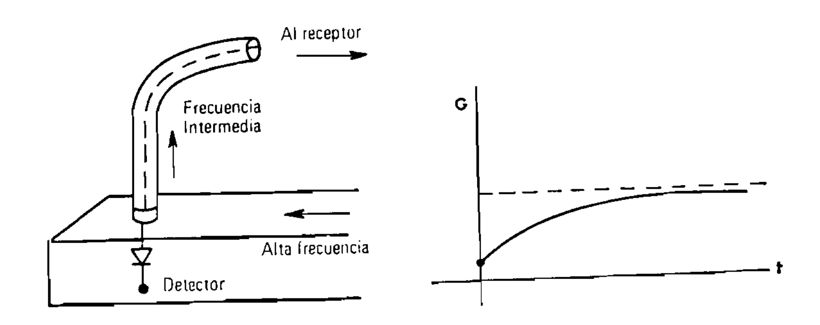


Fig. 18-7. Fig. 18-8.

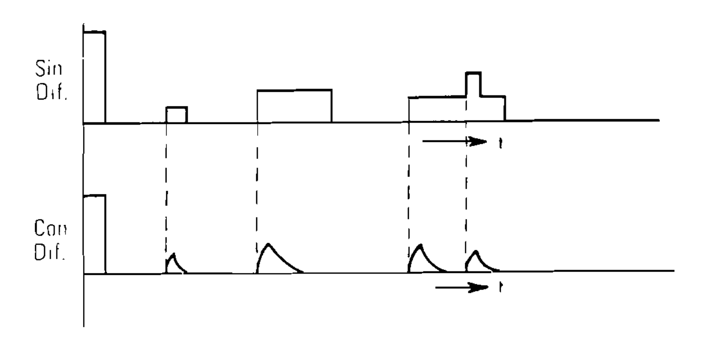


Fig. 18-9

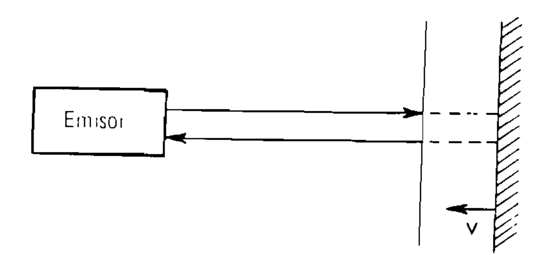


Fig. 18-10.

El efecto Doppler es la desviación de frecuencia debida al cambio de camino entre el transmisor y el receptor. Si, en la figura 18-10, el blanco se mueve a la velocidad v (acercándose) en un período 1/f el blanco se ha movido v/f (f, frecuencia del emisor). El tiempo de recorrido de la onda entre el emisor y el blanco y vuelta varia de un período al siguiente en la cantidad 2v/fc, siendo c la velocidad de propagación. El periodo de la onda emitida es 1/f y el de la recibida será:

$$\frac{1}{f}-\frac{2v}{fc}=\frac{1}{f}(1-\frac{2v}{c})$$

Como 2v/c es pequeño frente a la unidad, se puede decir que la frecuencia varía la cantidad 2vf/c aumentando si el blanco se acerca y disminuyendo si se aleja.

El radar Doppler más sencillo emite en onda continua y mezcla la onda que vuelve con una fracción de la del emisor, obteniendo así la frecuencia diferencia, proporcional a v. Midiendo esta frecuencia se tiene la velocidad (figura 18-11). El radar Doppler fue el primero que se usó para detectar la presencia de aviones enemigos. Actualmente se usa el esquema de la figura 18-11 para medir la velocidad de los vehículos (policía, por ejemplo) y la de aterrizaje de los aviones. Este tipo de radar sencillo (con mezcla directa, sin oscilador local) se denomina **homodino.**

La figura 18-12 representa un radar Doppler algo más complicado, llamado **heterodino** como el que se usa en localización de aviones y satélites. En el mezclador aparece la frecuencia del emisor f y la recibida f (1 + 2v/c). La mezcla de ambas frecuencias equivale a una modulación (en amplitud y frecuencia) de la primera por diferencia con la segunda. A continuación se separa esta modulación mediante un receptor superheterodino. A la salida se conecta un discriminador que regula la frecuencia del oscilador local para que la intermedia sea la deseada. Como antes, la medida de frecuencia de desviación se transforma fácilmente en medida de velocidad.

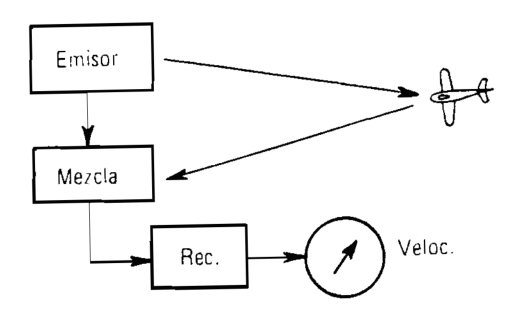


Fig. 18-11.

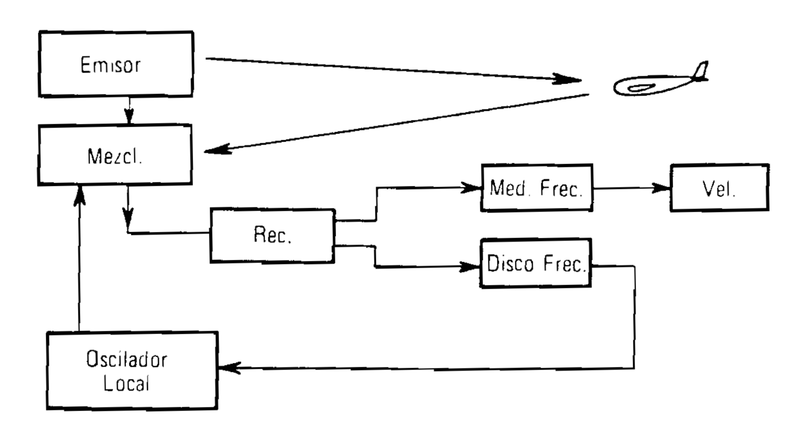


Fig. 18-12.

CAPITULO 19

CONTROL AUTOMATICO

19.1.- El problema del control.

El problema del control puede definirse de la siguiente manera (figura 19-1):

Un sistema X está caracterizado por una serie de variables x_i del tiempo ligadas entre sí y con otras por unas ecuaciones de ligadura. Las variables x_i se llaman **de estado** y componen un vector de n dimensiones que es el vector de estado del sistema. En las ecuaciones del sistema figuran las variables x_i, y, además, las u_j y p_k que definiremos a continuación. La variable independiente puede ser el tiempo t, en forma de variable continua o de intervalo n-simo de medida. En el primer caso (control continuo), las ecuaciones serán diferenciales y en el segundo (control discreto) en diferencias.

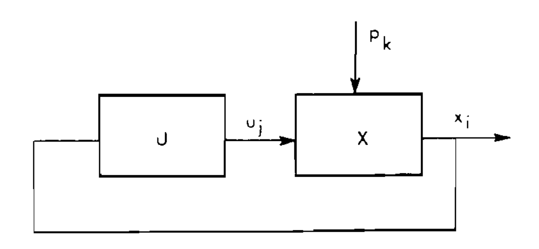


Fig. 19-1.

El sistema X depende de las variables de control u_j que forma el vector **u** de control. u_j son entradas del sistema; la dependencia no es reversible, es decir, x_i dependen de u_j pero no a la inversa. Las variables u_j son disponibles para nosotros; podemos fijarlas directamente o a través de ligaduras de otro sistema, el sistema de control.

Además, en el sistema X hay otra entrada indeseada, llamada perturbación **p** en forma de una serie de variables p_k que no están a nuestra disposición, son un ruido indeseado e imprevisible, aunque también se puede hacer algún tipo de pronóstico limitado sobre él.

El sistema X tiene una salida **observable z** que puede ser función de **x**, o ser directamente **x**. Para simplificar, podemos suponer que es directamente **x**, es decir que las variables de estado son observables.

La labor de conseguir un fin determinado, manipulando **u**, la hace un sistema **controlador U**. Este sistema utiliza los datos disponibles del problema, en general x_i, para fijar los u_j de forma que se consiga optimar el resultado. El resultado puede definirse como una función de las x_i que hay que optimar; en algunos casos será hacerla mínima o máxima, o bien que cumpla alguna condición determinada.

El problema del control es la síntesis del sistema **U**. Sin embargo, como ocurre frecuentemente en ingeniería, la solución se adopta en etapas sucesivas y no es corriente que se plantee en la forma que lo hemos hecho, general y brillante pero poco práctica. Por eso a continuación pasaremos a exponer el problema en forma mucho más restringida, que es como lo vamos a abordar aquí.

19.2.- Problema restringido.

Supongamos que el sistema tenga una sola variable de estado x(t), que esté sometida a una ecuación diferencial que la ligue con la variable de control u(t) y la perturbación p(t). El problema consiste en encontrar un controlador U que nos dé la relación u(t) como función de x(t) para que se cumpla una condición determinada. Esta condición en nuestro problema restringido es que el **error** o diferencia entre el valor "actual" de x(t) y el valor "deseado" $x_0(t)$ sea lo más pequeño posible. En la práctica es corriente que u(t) sea función de x(t) y $x_0(t)$ a través del error $x - x_0$, tal y como se representa en el diagrama de la figura 19-2, que consta de un **comparador** que efectúa la medida del error x_0-x, un **controlador** U y un **sistema** X. Una versión elemental del controlador U es que es un aparato que "observa" el error x_0-x y actúa sobre u (el mando del sistema) "para anularlo". Si se supone que la actuación de u sobre x es normalmente positiva (dx/du > 0), lo lógico es hacer crecer u cuando x ha disminuido (para que vuelva a aumentar),que es por lo que el error se mide como se indica en la figura, x_0-x, con una inversión en el signo de x (suponiendo que U actúa según el signo de x_0-x).

Este cambio de signo en el comparador, que es distinto al convenio que se hacia en otro capítulo sobre realimentación, es convencional en la teoría del control.

En este planteamiento sencillo del problema del control, al

que nos referiremos con la denominación más tecnológica de "regulación automática" suponemos además que el control es continuo (actúa en todo momento) y que las ecuaciones de X y de U son diferenciales lineales de coeficientes constantes.

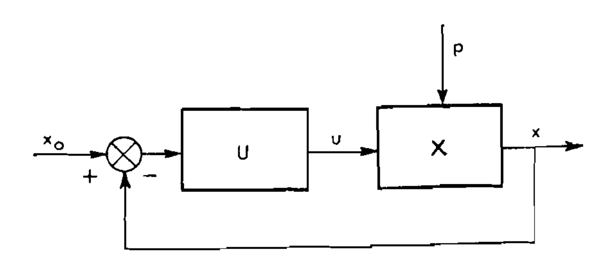


Fig. 19-2.

A primera vista pudiera parecer que las restricciones a que hemos sometido el problema son tales que va a perder aplicación práctica. Nada más lejos de la realidad; en la técnica de aplicación diaria, la tendencia a simplificar los problemas, en un esfuerzo por distinguir lo esencial de lo accesorio, es muy grande. Casi todos los problemas técnicos se agrupan en los dos grandes extremos de "lineales" o de "conmutación". Los problemas no lineales suelen tratarse como problemas lineales en los cuales se introducen variaciones posteriores de los parámetros. Los métodos de ataque directo de los problemas no lineales suelen quedarse en especulaciones teóricas de poca aplicación práctica. Ellos son el llamado "vacío" (gap) entre los teóricos y los prácticos que constantemente preocupa a los técnicos de todo el mundo, como viene reflejándose en congresos y reuniones internacionales, que se esfuerzan por rellenar ese vacío sin conseguirlo.

Aclaremos también que la palabra "sistema" tiene varios significados posibles, según el nivel de generalidad que se le dé:

a) Puede significar el conjunto de los subsistemas X y U (figura 19-2), dándoles a éstos la denominación de subsistemas.

b) Puede significar cualquiera de las partes X y U, habiendo entonces un sistema XU que comprende los sistemas X y U.

c) También pueden llamarse sistemas a partes de X o U que tengan más de una componente.

Nosotros la utilizaremos en cualquier acepción de las citadas, procurando precisar en cada caso el tipo de sistema que nos referimos.

19.3.- Ejemplos de sistemas controlados. Modelos.

Casi todos los equipos técnicos tienen algún "mando" o control (acepción restringida) que, de alguna forma, influye

sobre la variable de estado. Así, por ejemplo, el acelerador de un motor de explosión es un mando cuya posición influye sobre las revoluciones por minuto. Normalmente hay otros factores que influyen sobre la variable de estado, por ejemplo la carga del motor. Así el motor como sistema puede representarse (figura 19-3) como un bloque con una **entrada** o mando, que es el acelerador, y cuya posición está definida por el ángulo de la mariposa, θ . La carga del motor o par resistente T es una perturbación (puede considerarse como tal por influir indeseadamente en las revoluciones por minuto) y la salida está representada por las r.p.m., n, que será función de θ y T. La ecuación del bloque es la que liga n con θ y T. θ es la variable de control y n la de estado. El motor es regulable si tiene influencia suficiente para poder variar n dentro del margen de valores que deseemos cualquiera que sea T (dentro, también, de las limitaciones previsibles).

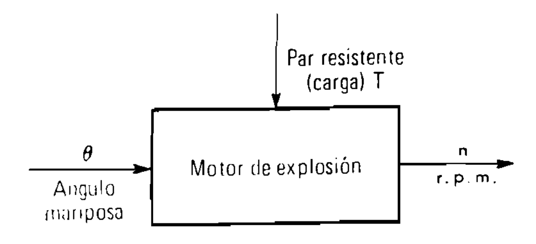


Fig. 19-3.

Del mismo modo, un motor eléctrico de c.c. con inductor a corriente constante y controlado por la tensión aplicada al inducido es un sistema cuya variable de entrada es la tensión eléctrica aplicada, la variable de salida, la velocidad (o el ángulo de posición del eje) y la perturbación la carga. Para estudiar la influencia de otros factores, puede suponerse algún otro tipo de perturbación, como, por ejemplo, la variación de corriente en el inductor (que debería ser constante), etc.

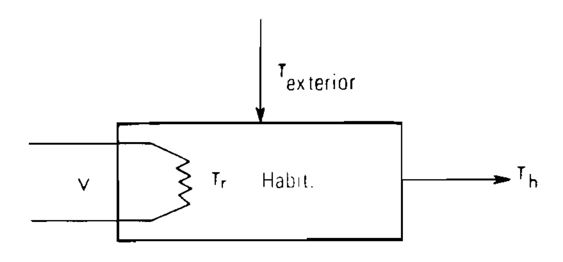


Fig. 19-4.

En los sistemas más complicados es necesario a veces bastante experiencia para seleccionar las variables adecuadas. Por ejemplo, supongamos que se trata de la calefacción de una habitación mediante un radiador eléctrico (figura 19-4). La variable de salida es la temperatura de la habitación T_h. Como

variable de entrada puede tomarse la temperatura del radiador, en cuyo caso el radiador mismo está excluido del sistema (e incluido en el regulador) o bien la tensión eléctrica que se aplica al radiador (con lo que éste queda incluido en el sistema). Las perturbaciones son diversas: apertura de ventanas, temperatura exterior, número de personas en la habitación (que calientan con su temperatura corporal), etc. Normalmente se estudia su efecto por separado, considerando una perturbación y dejando constantes los demás factores que influyen. Por ejemplo, la temperatura exterior.

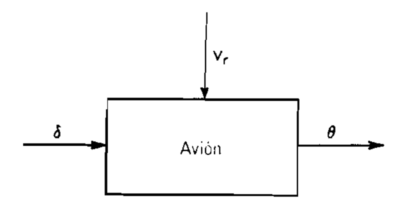


Fig. 19-5.

Algunas veces se estudian aspectos parciales de un problema, cuando las variables son separables. Por ejemplo, en el control de un avión, se estudian los mandos longitudinales y los laterales por separado. Considerando el vuelo "simétrico" del avión (figura 19-5), la variable de salida es la posición del eje del avión respecto a la horizontal, llamado "ángulo de asiento" θ. Como variable de entrada puede tomarse la posición angular del mando del timón de profundidad δ. Las perturbaciones son diversas: cambios de velocidad, ráfagas verticales, cambio en la potencia del motor, etc. Normalmente se considera una a la vez; por ejemplo la velocidad de ráfaga vertical V_r. Las ecuaciones de mecánica del vuelo resuelven el problema de encontrar θ en función de δ y V_r.

En todos estos ejemplos debe quedar claro que el sistema es conocido matemáticamente cuando podemos escribir la relación entre la salida y las entradas (control y perturbación) en forma de ecuación diferencial. Conociendo las entradas en función del tiempo y los valores iniciales de la salida y sus derivadas podremos calcular ésta en función del tiempo. El bloque que responde en la forma que nosotros establecemos así (con las simplificaciones y supuestos que introducimos se llama **modelo**. El modelo es más fiel a la realidad cuantas menos simplificaciones hayamos hecho para llegar a él. Todas nuestras deducciones posteriores están ya sujetas a la precisión de nuestro modelo. Un perfeccionamiento típico de la teoría del control, a introducir a posteriori, es la comprobación de los errores a que nos han conducido los errores del modelo. Pero la primera impresión del sistema, la que normalmente sirve para decidir el tipo de control que usamos, es la que obtenemos a través del modelo.

19.4.- Tipos de control.

El uso del control para obtener una respuesta del sistema puede hacerse de las siguientes formas:

a) Previendo o calculando el resultado de cada maniobra y ejecutando las maniobras que se crea que van a dar el resultado apetecido. Pueden incluso programarse para que el resultado sea lo mejor posible.

Por ejemplo, si se quiere calentar una habitación puede encenderse la calefacción al máximo durante dos horas por las mañanas y a media potencia el resto del día. O si se quiere mover una bomba con el motor de explosión, puede ponerse la mariposa a 2/3 del recorrido.

Esta forma de regular se llama en circuito abierto. Es obvio que se trata de una forma poco elaborada y no la consideramos automática.

b) Programando o actuando sobre el mando de acuerdo con la perturbación que se pueda medir o prever.

Por ejemplo, en la calefacción de una habitación puede darse al radiador eléctrico una tensión decreciente con la temperatura exterior según una dependencia que nosotros juzguemos conveniente para obtener la temperatura interior deseada. O al mover una bomba con el motor de explosión actuar sobre la mariposa según el flujo que sea necesario bombear.

Esta forma se llama en circuito abierto con introducción de la perturbación. Es ya una regulación automática; si el modelo es suficientemente bueno corregirá en un cierto margen la influencia de la perturbación previsible aunque nunca la de la perturbación imprevisible o ruido. Su estudio no tiene dificultad especial por lo que no lo estudiaremos por separado (figura 19-6).

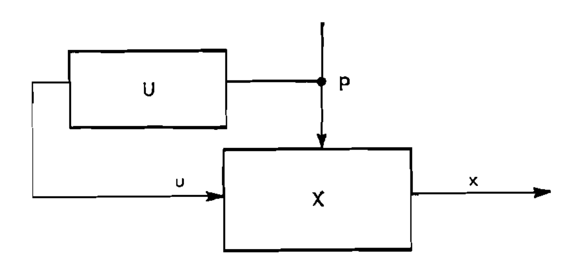


Fig. 19-6.

c) Incluyendo una información sobre el resultado para actuar sobre el mando (figura 19-7).

Por ejemplo, en la calefacción eléctrica de la figura 19-4, haciendo que V sea función decreciente de T_h. Cuando la habitación se calienta, V disminuye y deja de calentarse. Si se

enfría (por cualquier motivo, previsible o imprevisible), V aumenta y calienta más, compensando así el enfriamiento.

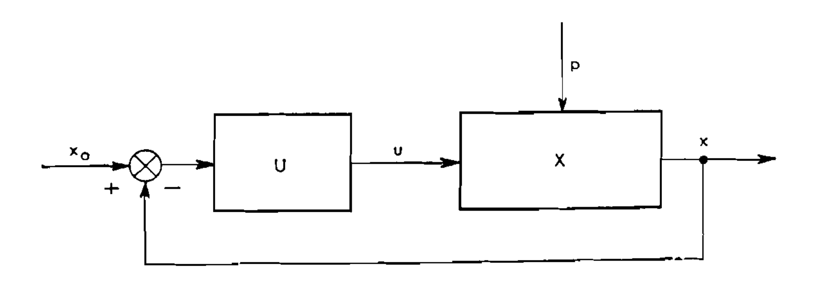


Fig. 19-7.

Este tipo de regulación se llama en circuito cerrado o con realimentación (la realimentación fue definida ya anteriormente y consiste en "llevar" a la entrada una función de la salida), y es el que propiamente constituye la regulación automática y será objeto de estudio en el resto de este capitulo y siguientes.

d) Un perfeccionamiento consiste en incluir en la regulación la parte previsible o medible de la perturbación (figura 19-8), cosa que facilita la regulación ya que "descarga" en parte al regulador automático que no tiene que esperar al efecto de la perturbación para actuar en contra.

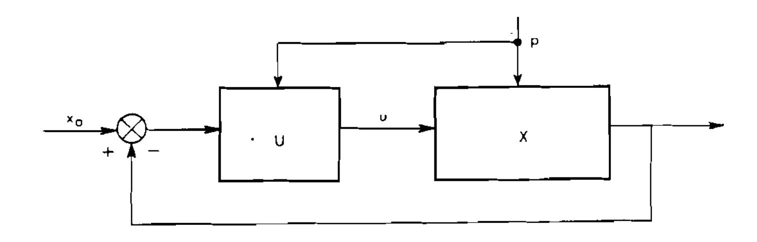


Fig. 19-8.

19.5.- Ejemplos de controlador.

La parte del sistema que efectúa la labor lógica de calcular la maniobra necesaria para el fin deseado y manipula el mando en consecuencia se llama controlador. El controlador U parte de la medida efectuada sobre la variable de estado (a través de la realimentación) y del valor deseado de esta variable, o bien de la diferencia o desviación, y da un valor a la variable de control u (figura 19-2).

A veces se utiliza el vocablo "regulador" como sinónimo de controlador y, otras veces, se reserva la palabra regulador para el caso de que se desee mantener x constante, es decir, x_0 constante o cero (figura 19-2). Aquí lo utilizaremos como sinónimo, ya que en la práctica se hace también así.

Un primer ejemplo de controlador rico en aplicaciones es el operador humano. En todos los ejemplos dados con anterioridad puede introducirse el control manual con observación del resultado. Por ejemplo en el caso del motor de explosión de la figura 19-3, el controlador será el operario que maneja el acelerador mirando al instrumento que marca el número de revoluciones (figura 19-9). Del mismo modo puede ver la temperatura de la habitación y "regular" la calefacción de la figura 19-4 o bien darse cuenta de su ángulo de asiento (mirar al horizonte) y pilotar el avión de la figura 19-5.

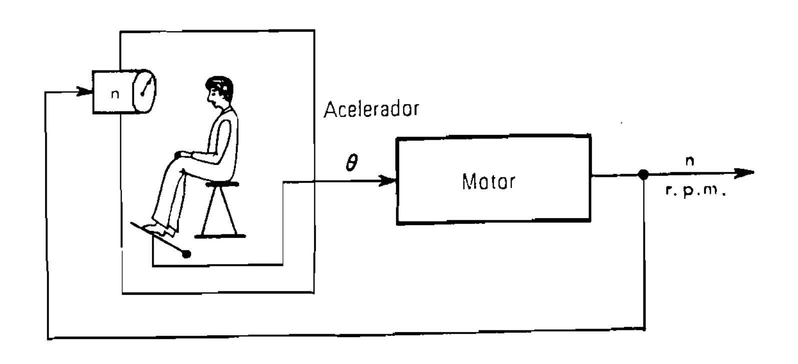


Fig. 19-9.

El operador humano es el antecedente histórico de los controladores automáticos que han ido sustituyéndole en las labores más fáciles, primero, e incluso en las complicadas después. El controlador (humano o no) puede subdividirse en varios subsistemas que constituyen tres partes:

a) Sección de **medida**, que efectúa la medida de la variable de estado. En el caso del operador humano, ojos, etc. En otro caso, termómetros, manómetros, horizontes artificiales, etc.

b) Sección **lógica**, que hace los cálculos necesarios para tomar medidas que corrijan el error. Es el cerebro, humano, electrónico, etc., del sistema.

c) Sección **actuadora**, que acciona el mando. Son los músculos del controlador o bien motorcitos eléctricos, actuadores hidráulicos, etc., que aportan la potencia necesaria para mover el mando del sistema controlado.

La medida del error es una operación lógica; a veces la sección de medida da ya el error y otras sólo da la medida de x, siendo necesario incluir en la parte lógica la operación x_0-x. En los esquemas generalmente se representa el comparador por separado (figura 19-2), pero es necesario observar que en la práctica el aparato de medida puede incluirse en el sistema comparador, en un comparador separado o en el controlador y que la abstracción de considerar el comparador separado es puramente

teórica.

La parte lógica es normalmente electrónica, salvo en controladores elementales en los que basta con utilizar directamente la medida (sin manipulaciones). Los actuadores son dispositivos que están adaptados a la naturaleza de la variable de control. Generalmente son motores, pero también pueden ser otros elementos.

Tradicionalmente se llamaba al controlador "servomecanismo" haciendo referencia al carácter normalmente mecánico del actuador, y se solía decir que podían clasificarse en eléctricos, hidráulicos y neumáticos, según la naturaleza de la fuente de energía que se usaba. Hoy día ha perdido importancia esta clasificación, ya que el estudio de los sistemas automáticos se hace más con referencia a su lógica, prescindiendo de la naturaleza de los componentes. Sin embargo, sigue siendo válido que el actuador puede ser de una de esas tres clases, según la fuente de la energía que utilice.

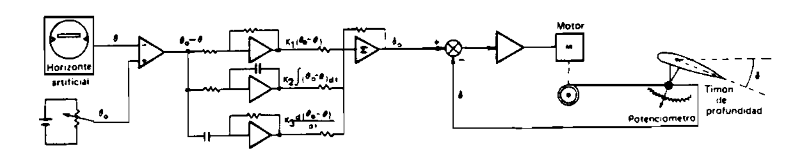


Fig. 19-10.

Un ejemplo de controlador puede ser el piloto automático longitudinal, que es el que puede sustituir al piloto humano en el control de la variable simétrica del avión. la figura 19-10 es un diagrama del piloto automático (longitudinal). La sección de medida será un horizonte artificial que mide el ángulo θ. Esta componente está basada en el giróscopo y es un transductor de posición, ya descrito en el capítulo correspondiente. El comparador es un amplificador diferencial que resta el valor actual de θ del deseado θ_0 que fija el piloto mediante un potenciómetro. (El piloto humano ya no va a mantener θ constante e igual a θ_0, como haría sin regulador, sino a fijar θ_0, con lo cual puede hacer que el avión vuele horizontal o baje o suba en régimen de regulación automática).

A continuación del comparador viene la sección lógica que hace las operaciones necesarias según el error observado. La separación de comparador y parte lógica o "amplificador" es una abstracción; obsérvese que la resta y el cambio de signo se hace ya en un amplificador analógico que en realidad forma parte del resto del calculador electrónico. No obstante, para el tratamiento en capítulos sucesivos, supondremos que primero se hace la operación $\theta_0 - \theta$ (comparador) y luego se amplifica. A la sección lógica se la llama en la práctica **amplificador**, a pesar de que no es sólo un amplificador, sino todo un calculador analógico o digital. Por ejemplo, podemos usar un control del tipo "PID" (proporcional-integral-diferencial), en que se haga

una corrección **"proporcional"** cuando haya una desviación de θ, una corrección **"diferencial"** cuando haya una velocidad de desviación, que nos diga que θ tiende a desviarse más y una corrección **"integral"** cuando haya una pequeña desviación de θ que dure mucho tiempo. Detrás de la sección lógica vendrá el **actuador**. En este caso podemos suponer que sea un motor eléctrico que actúe, a través de un reductor, sobre un tambor. En este tambor llamado **"capstan"** se enrolla el cable que va a actuar sobre el timón de profundidad, que es así accionado por el motor. La dinámica del motor (una integración), debe estar incluida en la parte lógica, de forma que no se haga. una integración más. Esto se consigue mediante una realimentación adicional, con lo que el motor constituye un sistema completo de regulación incluido en el sistema total. Este servomecanismo final hace que δ siga a δ_0 que es el valor deseado.

El timón de profundidad es el mando del avión, δ es la variable de control del avión. Las ecuaciones de vuelo serán las que determinen la respuesta de θ a δ, cerrando así el circuito a través del avión, sistema controlado.

19.6.- Realimentación.

La utilización de la salida para valorar el error a la entrada se llama **realimentación**. De la realimentación ya hemos hablado en otro capítulo, en el que suponíamos que a la entrada se le sumaba una fracción β de la salida. En los ejemplos que hemos visto hasta ahora en este capitulo hemos supuesto $\beta = -1$, caso que se denomina realimentación unitaria. El motivo por el que se resta ya lo hemos indicado, es para que la regulación actúe en contra del error; de ahí el signo menos.

El sistema de circuito cerrado funciona tratando de anular el error a la entrada, dado por el comparador. De ahí que, si se quiere anular el valor $x_0 - x$ (figura 19-2), lo lógico sea utilizar realimentación unitaria. Sin embargo, esto no es normalmente posible, ya que lo que se lleva al comparador (ver los ejemplos del párrafo anterior) no es x sino el valor medido de x que, hasta ahora hemos supuesto que coincide con x.

Si el valor medido de x, sea z, no coincide con x sino que hay un "coeficiente" de medida β, tal que $z = \beta x$, el esquema del circuito cerrado de la figura 19-2 se transforma en el de la figura 19-11. El regulador o controlador U actúa siempre que x_0-z no sea cero, tratando de hacerlo nulo. Con ello se tiende a que (si el controlador funciona correctamente) sea $x_0 - \beta x = 0$, es decir $x = x_0/\beta$. Esto se suele expresar diciendo que **el circuito cerrado efectúa la operación inversa de la realimentación.** Si $\beta = 1$ hace que $x = x_0$, caso particular. En principio este resultado no debe sorprendernos, es el que obtuvimos en el capítulo de realimentación; la diferencia es que ahora vemos β como algo indeseado, coeficiente de medida inevitable que nos aleja de nuestro ideal de que tome el valor deseado x_0. El resultado que obtenemos no es $x = x_0$, sino $x = x_0/\beta$ por muy bueno que sea el controlador. La regulación automática, tal y como la hemos estudiado, puede llegar a ser todo lo buena que sea la medida,

pero no más. Por eso es esencial que la medida sea fiel. Claro que se puede contar con este efecto y hacer que el valor deseado que se introduce en el comparador no sea el deseado para la variable de estado x_0, sino que sea βx_0, **siempre que β sea conocido.**

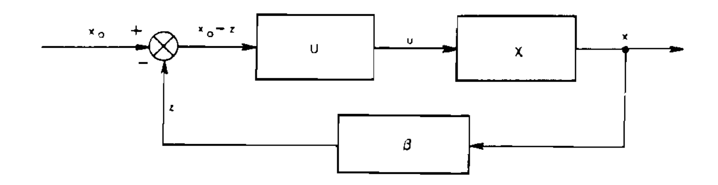


Fig. 19-11.

En los ejemplos del párrafo anterior, β estaba incluido en el controlador; para su estudio será necesario separarlo, para llevarlo al esquema de la figura 19-11. Así en el ejemplo del motor, β es la transformación que hace el tacómetro (la realimentación es el tacómetro), en el de la calefacción es el termómetro y en el del piloto es el horizonte artificial. Los controladores automáticos actúan para que las revoluciones dadas por el tacómetro, la temperatura dada por el termómetro o el ángulo de asiento dado por el horizonte artificial sean constantes.

19.7.- Perturbaciones.

Se llaman perturbaciones a aquellas variables que intervienen en el sistema sin formar propiamente parte del mismo. Dentro de esta definición todavía hay dos clases:

a) Perturbaciones medibles, como por ejemplo, la temperatura exterior a una habitación cuya calefacción se quiere regular. Se puede medir con un termómetro exterior.

b) Perturbaciones no medibles, como por ejemplo, la temperatura de la gente que entra en la habitación. Bien entendido, que se trata de un concepto práctico ya que, con todo rigor, se podría medir la temperatura de las personas que entran en la habitación, pero no es pensable una regulación automática de la calefacción que tuviese que hacerlo en la realidad.

En este concepto pueden incluirse también todos los errores de los demás parámetros y variables de la regulación, que no son medibles a efectos de la regulación misma.

También se incluyen los ruidos de todas clases, blancos o gaussianos que intervienen en el sistema.

Las perturbaciones del tipo a) no son auténticas perturbaciones ya que podrían clasificarse formalmente como simples parámetros del sistema, junto con las variables cuyo valor es previsible. En la práctica se hace lo contrario, es

decir, se clasifican variables previsibles como perturbaciones, por simple comodidad. Así, por ejemplo, la temperatura exterior es parcialmente previsible (recuérdese que cambia de invierno a verano y de día a noche). La diferencia de tratamiento técnico entre previsible y no previsible es:

a) Si se prevé el cambio de temperatura, se hace un horario o **"programa"** de calefacción y se cambian los parámetros del sistema para que la dinámica del día a la noche y de invierno a verano sea distinta. En todo caso es necesario contar con una variación de temperatura no prevista que actúa por separado, que pasa a ser la verdadera perturbación.

b) Se puede considerar toda la variación de temperatura exterior como no prevista, medir la temperatura e introducir cambios en el controlador de acuerdo con las necesidades (figura 19-7). Esta técnica se llama "introducción de la perturbación" y tiene la ventaja de que ya cuenta con la variación de temperatura, prevista o no, en su totalidad. Más adelante insistiremos sobre la ventaja de esta técnica.

Las perturbaciones no medibles y las aleatorias influyen sobre el sistema y no es posible tenerlas en cuenta en el controlador. En realidad son la razón de ser de la regulación automática, como veremos después.

19.8.- Reguladores PID.

Una vez vistas las partes del sistema de regulación automática, conviene hacerse una idea cualitativa del funcionamiento del conjunto, sin perjuicio de que más adelante se estudie cuantitativamente el comportamiento dinámico del sistema. Ya en varios párrafos hemos hablado del concepto elemental de "regular contra el error", pero ahora lo vamos a precisar un poco más. Nos referiremos exclusivamente a la regulación en circuito cerrado, única que es verdaderamente "automática", y vamos a describir el comportamiento típico de un regulador PID.

En primer lugar aclaremos que no sería necesaria la regulación si no fuese por la presencia de perturbaciones. El circuito cerrado es un verdadero mecanismo de lucha contra el error producido por la perturbación. Podemos decir que la realimentación "nota" el efecto de la perturbación ;algo así como que, ya que no se mide la perturbación sí se mide su efecto (el error). La medida de las r.p.m., la temperatura de la habitación o el ángulo de asiento del avión en los ejemplos de las figuras 19-3, 19-4 y 19-5, es una medida que "ve" el efecto producido por nuestra regulación y por la perturbación.

Una vez medido este "resultado" y valorado el error (restando del valor deseado, véase la figura 19-2, comparador) se inserta un controlador U que actúe sobre la variable de control para tratar de anular el error. Cuando la corrección se tome proporcional al error observado, la regulación se llama **proporcional (P)**. El inconveniente de la regulación proporcional

es que para que haya corrección tiene que haber un error que, por consiguiente, no llega a anularse; queda un error residual. Tomemos como ejemplo la regulación de temperatura de la figura 19-12, con la habitación de la figura 19-4, que se desea mantener a la temperatura T_0. Si, debido al frío ($T_{ext.}$ bajo), desciende T_h, T_r aumenta proporcionalmente (si el regulador es P) al valor $T_0 - T_h$, con una cierta amplificación. El error $T_0 - T_h$ disminuirá, llegándose a un equilibrio cuando la temperatura T_r sea la necesaria para mantener $T_0 - T_h$ pequeño (pero no nulo), podría pensarse que basta una desviación $T_0 - T_h$ de un grado para aumentar T_r a 100 grados, manteniendo así el error (si esa calefacción es suficiente) en menos de un grado.

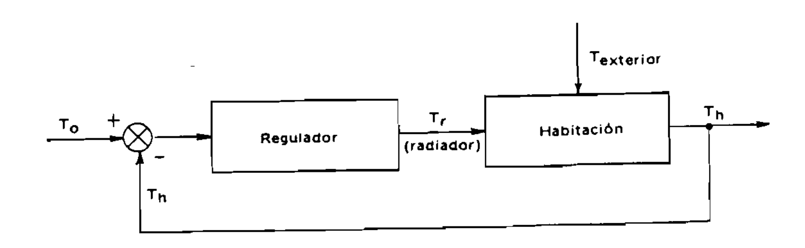


Fig. 19-12.

El regulador puede hacer algo más. Si nota que el error persiste mucho tiempo, puede tomar correcciones cada vez mayores, tratando de eliminar la parte residual. A esto se le llama **corrección integral (I)** y se hace tomando una desviación de la variable de control proporcional a la integral del error. Así, un error pequeño pero permanente tiende a desaparecer con el tiempo.

Por último podemos fijarnos en la velocidad de reacción del sistema ante una perturbación brusca. Imaginemos que, de repente, cambia la temperatura exterior (sale el sol, etc.). En seguida empieza a calentarse la habitación y, a través de la regulación proporcional, empieza a disminuir la temperatura del radiador. Es fácil pensar que, si la temperatura aumenta constantemente, es mejor disminuir más la temperatura del radiador, adelantándose a los acontecimientos, en forma proporcional a la velocidad de cambio. A esto se le llama **corrección diferencial (D)** y es un perfeccionamiento de la dinámica del sistema.

En las tres técnicas de regulación (**P, I, D**) que hemos visto, hemos esperado a ver el efecto de la perturbación para tomar medidas en contra. Evidentemente esto es una degradación de la información; por eso, cuando podemos medir la perturbación, es mejor anticiparse y corregir para neutralizarla en lugar de esperar a su efecto.. En el ejemplo de la figura 19-12, es lógico calentar más cuando se observe un descenso de $T_{ext.}$, liberando así al circuito cerrado de tener que luchar contra ese efecto. A este perfeccionamiento se le llama "**introducción de la perturbación**".

19.9.- Precisión y estabilidad.

Toda la descripción que hemos hecho hasta ahora en el párrafo anterior tiende a obtener una mayor precisión en la regulación. Sin embargo, es preciso conocer la limitación natural de esta tendencia, que está constituida por la estabilidad. A primera vista parecería que se podía hacer crecer ilimitadamente la amplitud de la "corrección" y conseguir más y más precisión. Sin embargo, la combinación de amplificación y retardo (hay que contar con retardos en la medida, retardos en el control, etc.) siempre conduce a inestabilidad, verdadera catástrofe del sistema de regulación automática.

Para comprender la inestabilidad podemos poner el ejemplo de la persona que se ducha mezclando agua fría y caliente y trata de controlar la temperatura de la mezcla manejando el grifo de agua caliente. Si nota que la ducha está fría abre más el grifo de agua caliente, y si nota que está demasiado caliente lo cierra.

Sin embargo, debido al tiempo que tarda el agua en recorrer la cañería entre el grifo y la ducha, puede ocurrirle al duchante nervioso el siguiente desagradable efecto: Al notar que tiene frío abre el agua caliente; como todavía no nota nada, abre más y más hasta el final. De repente, y pasados unos instantes, se abrasa al recibir toda el agua caliente que había abierto antes. Al notarse quemado, cierra el agua caliente (hasta el final, ya que no nota ningún efecto instantáneo); pasados unos instantes recibe una ducha helada, por lo que abre el agua caliente, etc....

El duchante avezado puede estabilizarse. Abre cautamente el grifo y aguanta unos instantes, anticipándose al efecto y poniéndose muy contento cuando llega. En su momento veremos que lo que hace es estabilizar por avance de fase.

Lo que queremos aclarar ahora es que hay un dilema **estabilidad-precisión** y que su solución equilibrada va a ser el verdadero punto clave del éxito de la regulación.

19.10.- Proyecto de regulación automática.

El proyecto de una regulación automática suele hacerse por etapas sucesivas de síntesis-análisis, como es frecuente en ingeniería. No es corriente que a la vista del sistema a controlar se pueda "sintetizar" un controlador adecuado. Lo normal es que se elija uno que la experiencia ha aconsejado en problemas análogos y se analice, para probar si cumple las especificaciones exigidas. La regulación P I D que hemos visto antes es sólo un ejemplo al que frecuentemente se recurre en la experiencia, con las salvedades correspondientes a cada caso. Por ejemplo, en el caso de que el sistema a controlar incluya ya una integración, hay que contar con ésta en el conjunto ya que, entonces, la parte P del regulador se convierte en I, la D en P y no se puede poner una parte en I (ya veremos en su día que dos integraciones conducen a inestabilidad). El controlador óptimo (el diseño óptimo) se obtiene por etapas sucesivas de análisis y mejora. El estudio de la regulación automática suele

comprender:

a) Propuesta de un controlador, análisis del comportamiento del sistema en cuanto a precisión, estabilidad y otras actuaciones o cualidades (respuesta de frecuencia, etc.).

b) Estudio de posible mejora de la dinámica, consiguiendo un comportamiento estable con mayor precisión. (A esto se le llama compensación).

c) Posible aplicación de técnicas de optimación, como las que se verán en el párrafo siguiente.

19.11.- Técnicas de optimación.

El diseño óptimo del controlador debe tener en cuenta las características del sistema, reflejadas lo más perfectamente posible en el modelo que utilicemos.

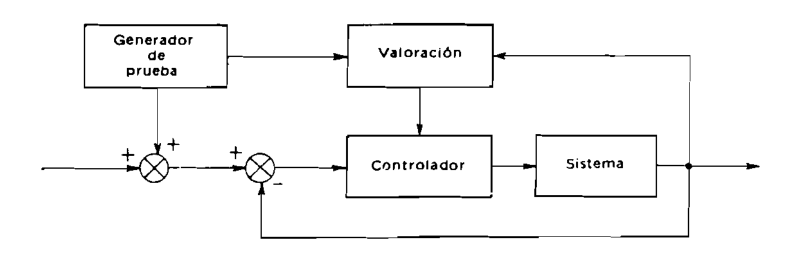


Fig. 19-13.

Se llama **optimación** a la técnica de mejorar la regulación utilizando información adicional de la realidad, bien por divergencias con el modelo, bien por errores de medida, ruidos etc.... En principio, estas técnicas pueden dividirse en dos clases:

a) **Técnicas de mejora de la medida** (o minimización de los errores).- Por redundancia de alguna clase (por ejemplo, utilizando dos canales para la misma medida) o por filtrado utilizando alguna característica de los errores que los haga detectables (por ejemplo, errores que aparezcan en forma sinusoidal etc...).

b) **Técnicas de aprendizaje.** Basadas en la variación de los parámetros del controlador, según el resultado obtenido en "maniobras" normalizadas o normales.

Las técnicas del tipo a) consisten en la inclusión de filtros. Cuando se trata de eliminar errores de características conocidas, como es el error sinusoidal inducido por la red eléctrica, se incluye un filtro normal que lo elimine. La

generalización de esta idea, aplicándola a errores estocásticos, pero de distribución estadística conocida (gaussiano, por ejemplo) da lugar a los llamados "filtros de Wiener y Kalman", cuyo tratamiento se sale del propósito elemental de este capítulo. También se aplica un "filtrado" al caso de tener que valorar varias fuentes de medida de la misma variable, tomando un valor combinado que viene dado por la evaluación más probable según la credibilidad de las fuentes, su frecuencia de medida, etc....

Las técnicas de aprendizaje son un control del control (figura 24-13); un generador de maniobra provoca una maniobra de prueba. La reacción del sistema se valora en un "comparador de valoración" y, según el resultado, se corrige el controlador.

CAPITULO 20

COMPONENTES Y SUBSISTEMAS DE LOS SISTEMAS DE CONTROL

20.1.- Normalización de la ecuación diferencial.

Cada bloque del sistema de control, sea un subsistema o un componente y sea correspondiente al sistema controlado, al controlador o a la realimentación, está caracterizado por una ecuación diferencial, que supondremos lineal de coeficientes constantes que relaciona, la salida con la entrada o las entradas. Supondremos que se trata de una sola entrada, ya que el caso de perturbación lo trataremos siempre como una variable que se suma en algún sitio (ver figura 20-1). El sistema que se controla se divide así en los subsistemas S_1 y S_2.

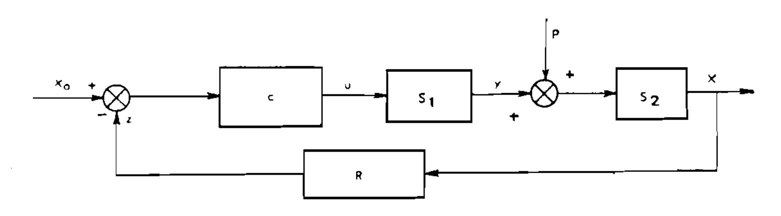


Fig. 20-1.

Cada bloque tiene una entrada x y una salida (fig. 20-2) y una ecuación diferencial del tipo:

$$a_0 \frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} + \ldots + a_n y + a_{n+1} = f(x)$$

donde se conocen los coeficientes $a_0, \ldots a_{n+1}$, $f(x)$ en función del

tiempo (o f(x) y además x(t)) y n valores iniciales de y.

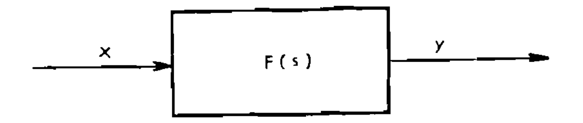


Fig.20-2.

En primer lugar normalizaremos la ecuación haciendo el cambio de variable $y = y' - a_{n+1}/a_n$ con lo cual la nueva ecuación en y' no tendrá término independiente ; el resto queda del mismo orden. Con este cambio de origen, en el caso particular del reposo, con todos los valores iniciales cero y f(x) = 0, el valor y = 0 es solución de la ecuación; equivale pues a tomar el "reposo" como valor origen de la y, por eso se llama **normalización de origen.**

Una segunda etapa de normalización (que no se hace siempre) es hacer el cambio de variable:

$$y=\frac{y'}{a_n}-\frac{a_{n+1}}{a_n}$$

Con este cambio se hace 1 el coeficiente de la y, y cambian todos los demás, quedando la ecuación de la forma:

$$A_0\frac{d^n y'}{dt^n}+A_1\frac{d^{n-1}y'}{dt^{n-1}}+\ldots+y'=f(x)$$

En esta nueva forma se consigue que si f(x) = C constante, la solución y = C sea solución estacionaria de la ecuación. Es decir que con $t \to \infty$, una vez amortiguados los transitorios (que son los que deben acoplarse a los valores iniciales), queda y = C = f(x). Por eso se le llama **normalización de escala**, ya que equivale a tomar la misma escala de medida para x y para y, que se corresponden en valor en régimen estacionario.

20.2.- Función de transferencia.

Sea

$$a_0\frac{d^n y}{dt^n}+\ldots+a_n y=b_0\frac{d^m x}{dt^m}+\ldots+b_m x$$

la ecuación de un bloque. Supongamos que partimos del reposo, es

decir, que todos los valores iniciales de y sus derivadas son cero.

Apliquemos la transformación de Laplace, llamando Y(s) a la transformada de y(t) y X(s) a la de x(t), (o, simplemente Y, X); recordando el teorema de derivación, con valor inicial cero:

$$a_0 s^n Y + a_1 s^{n-1} Y + \ldots + a_n Y = b_0 s^m X + b_1 s^{m-1} X + \ldots b_m X$$

de donde

$$\frac{Y}{X} = \frac{b_0 s^m + b_1 s^{m-1} + \ldots + b_m}{a_0 s^n + a_1 s^{n-1} + \ldots + a_n}$$

A este cociente se le llama función de transferencia F(s) del bloque y es independiente de X, dependiendo sólo de la ecuación diferencial del bloque y no de la entrada. La salida se puede despejar en la forma Y(s) = F(s) X(s).

La función de transferencia de un bloque F(s) es pues la relación de las transformadas de Laplace de la salida y de la entrada, con valores iniciales nulos, hipótesis que ya haremos casi siempre desde ahora.

Dada una entrada cualquiera x(t), para hallar la salida y(t) bastará hacer lo siguiente:

a) Hallar X(t) en unas tablas de transformación de Laplace.

b) Multiplicar por F(s) para hallar la salida Y(s) = F(s)X(s).

c) Buscar en las tablas la función y(t) que corresponde a Y(s).

La función de transferencia es pues un multiplicador en el plano de Laplace, es decir un "operador" que representa la transformación funcional que hace el bloque. (Siempre partiendo del reposo).

20.3.- Respuesta de frecuencia.

Sea un bloque como el de la figura 20-2 cuya función de transferencia es:

$$F(s) = \frac{b_0 s^m + \ldots + b_m}{a_0 s^n + \ldots + a_n}$$

y cuya ecuación diferencial era:

$$a_0 \frac{d^n y}{dt^n} + \ldots + a_n y = b_0 \frac{d^m x}{dt^m} + \ldots + b_m x$$

Si suponemos que tanto x como y son sinusoidales, podemos aplicar la transformación jω con lo que, recordando que las operaciones de derivar se sustituyen por multiplicar por jω, la ecuación en el plano de la pulsación, llamando ahora x e y a los complejos (fasores) que representan a x(t) e y(t), queda:

$$a_0 (j\omega)^n y + \ldots + a_n y = b_0 (j\omega)^m x + \ldots + b_m x$$

es decir, la **respuesta de frecuencia F(jω)** será:

$$\frac{y}{x} = \frac{b_0 (j\omega)^n + \ldots + b_n}{a_0 (j\omega)^m + \ldots + a_m}$$

y **se obtiene de F(s) sin más que cambiar s por jω.**

20.4.- Propiedades de F(s).

La definición de F(s) es la salida dividida por la entrada, es decir:

F(s) = Y(s)/ X(s)

y la propiedad más importante es que, como hemos visto, **la función de transferencia no depende de la entrada X(s).**

la respuesta de frecuencia se obtiene cambiando s por jω en F(s). También es válida la recíproca, la función de transferencia se obtiene de la respuesta de frecuencia cambiando jω por s.

Si en la fórmula Y(s) = F(s) X(s) hacemos X(s) = 1/s, es decir si damos una entrada x(t) igual al escalón unidad (que es la función cuya transformada de Laplace es 1/s) obtenemos la respuesta al escalón unidad Y(s) = F(s) / s, que es la función de trasferencia dividida por s. o, dicho al revés, **la función de transferencia es la respuesta al escalón unidad multiplicada por s.**

Vamos a dar a la entrada la función **percusión unidad (función de Dirac)**, es decir una función x(t) que vale cero excepto en el origen, no determinada en el origen y tal que su integral vale 1 a través del origen.

Su transformada de Laplace será X(s) = 1 de donde:

Y(s) = F(s) X(s) = F(s) y la función de transferencia es la respuesta a la percusión unidad (en el plano de Laplace).

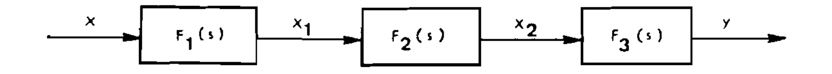


Fig. 20-3.

Por último aclaramos que si se conectan en serie varios sistemas de funciones de transferencia $F_1(s)$, $F_2(s)$ y $F_3(s)$ (tomando tres como ejemplo, figura 20-3), será:

$$Y(s) = F_3(s)X_2(s)$$
$$X_2(s) = F_2(s)X_1(s)$$
$$X_1(s) = F_1(s)X(s)$$

de donde, sustituyendo, se obtiene:

$$Y(s) = F_1(s)\ F_2(s)\ F_3(s)\ X(s)$$

Las funciones de transferencia de bloques conectados en serie se multiplican.

20.5.- Representaciones gráficas.

La función F(s) es una función de variable compleja y no admite representación en un plano; son dos funciones de dos variables y podrían representarse por dos superficies en tres dimensiones, cosa no suficientemente práctica.

En cambio $F(j\omega)$, la respuesta de frecuencia es una función compleja de variable real ω, equivale a dos funciones de una variable y sí se puede representar por dos curvas (parte real e imaginaria o módulo y argumento en función de ω. Generalmente se representa el módulo $|F|$ y el argumento, llamémoslo φ en función de ω y es la representación **directa de la respuesta de frecuencia e indirecta de F(s)** ya que a cada $F(j\omega)$ corresponde un F(s) sustituyendo $j\omega$ por s.

Hay varias formas de representar $F(j\omega)$, que son:

a) **Representación de Bode**, constituída por dos curvas que son $|F|$ y φ en función de ω. Constituye la respuesta de frecuencia que se ha estudiado en capítulos anteriores.

b) **Representación de Bode semilogarítmica**. En abscisas ω, en ordenadas en una curva log $|F|$ y en la otra φ. (Es decir, que sólo es semilogaritmica la representación en amplitud, quedando natural la de φ.

A veces se representa $|F|$ en decibelios lo que equivale a multiplicar lo anterior por 20, ya que $|F|$ dB = 20 log $|F|$.

c) **Representación de Bode logarítmica**, como la anterior pero en abscisas log φ.

d) **Representación de Black.** En una sola curva; en abscisas se representa φ y en ordenadas log $|F|$; sobre la curva se calibra ω como variable independiente, es decir que se elimina ω entre las funciones $|F|$ y φ para representar sólo $|F|$ (φ).

e) **Representación de Nyquist.** Es el diagrama en coordenadas polares de $F(\varphi)$. Es decir que para cada valor de ω se dibuja el vector **F**,ω en polares. El valor de ω se acota sobre la curva obtenida.

Nosotros **usaremos fundamentalmente el diagrama de Nyquist** y, algunas veces, el de Bode. Sin embargo, es necesario hacer notar que se podría construir la misma teoría sobre el de Black, como se hace en alguna literatura de este tema.

20.6.- Relación del diagrama de Nyquist con F(s).

Si ahora consideramos F(s) como una **transformación compleja** del plano **s al F**, los valores $F(j\omega)$ son los que corresponden en el plano F al eje $j\omega$ (imaginario positivo) en el plano s. De ahí se deduce que **el diagrama de Nyquist es el transformado del semieje positivo imaginario del plano s, tomando como transformación F(s)**.

El eje $-j\omega$, negativo imaginario, daría el diagrama conjugado, es decir, simétrico de el de Nyquist respecto al eje real.

20.7.- Sistemas de primer orden.

Son aquellos subsistemas cuya ecuación diferencial es de primer orden, del tipo $y'T + y = Kx$, donde se ha normalizado el origen aunque no la escala (debido a la constante K).

La función de transferencia será:

$$F(s)=\frac{K}{1+Ts}$$

La respuesta al escalón unidad 1/s será $Y(s) = K / (1 + Ts)$, que en el plano del tiempo es $y = K(1 - e^{-t/T})$, siendo la clásica respuesta"exponencial", de constante de tiempo T (figura 20-4). El valor final seria $y(\infty) = K$, llegándose en un tiempo T al valor $k(1-1/e)= 0'63$ K. En cinco constantes de tiempo se alcanza el valor 0'99 K, es decir casi el valor final. La constante de tiempo nos da una idea de la velocidad de reacción del sistema. La tangente en el origen corta a la asíntota en el valor T, la pendiente en el origen es K/T, como puede verse derivando en la expresión de y.

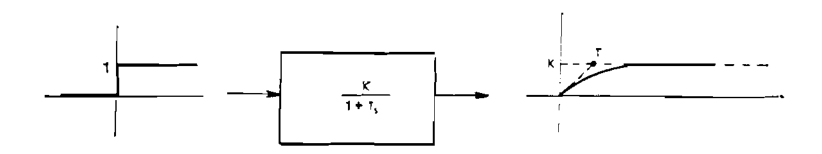


Fig. 20-4.

La respuesta de frecuencia será $F(j\omega) = K / (1 + j\omega T)$ correspondiente a la ya conocida de los sistemas RC. (Con RC = T). En la figura 20-5 está representada la respuesta de frecuencia (diagrama de Bode) con pulsación de corte $\omega_c = 1/T$ para la cual la respuesta vale 0,7 del máximo y la fase $-\pi/4$.

El diagrama de Bode logarítmico tiene la forma indicada en la figura 20-6. Para frecuencias bajas ($\omega << 1/T$) el valor es aproximadamente log K ya que $F(j\omega) = K$ si se desprecia $|\omega T|$ frente a la unidad. Para frecuencias altas ($\omega >> 1/T$) se puede despreciar la unidad; el módulo vale $K/\omega T$ y su logaritmo será ahora log K-log T-log ω, que en coordenadas logarítmicas es una recta de pendiente -1 y que corta a la otra (log K) cuando $\omega = 1/T$ (en cuyo caso log ω =-log T). Este punto corresponde a la pulsación de corte. (El punto $\omega = 0$ sería $-\infty$ en abscisas).

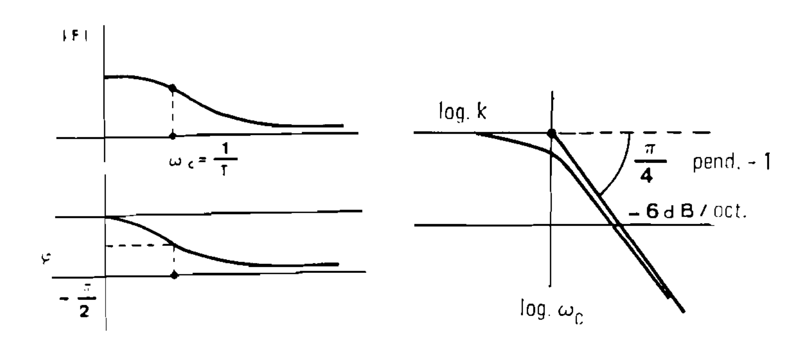


Fig. 20-5. Fig. 20-6.

La aproximación del diagrama por dos rectas, log K y otra con pendiente -1 a partir de $\omega = 1/T$ se llama **diagrama asintótico de Bode** y es válida lejos de ω_c, cuando ω es mucho menor o mucho mayor que la pulsación de corte 1/T.

La pendiente -1 en el diagrama logarítmico puede decirse que corresponde a decir que la amplitud se divide por dos cuando la frecuencia o pulsación se duplica. Dividir por 2 la amplitud se llama bajar 6 dB. Multiplicar por 2 la frecuencia se le llama "aumentar una octava" (procedente de la terminología musical). Por eso al descenso de la segunda asíntota de la figura 20-6 se le llama de "6 decibelios por octava".

Resumiendo, el diagrama asintótico de Bode, de un sistema

de primer orden, se compone de una horizontal a la altura log K y una recta que baja desde el punto de ruptura 1/T a razón de 6 dB por octava.

El **diagrama de Nyquist** es la representación de la curva transformada de $j\omega$ según la transformación compleja K/(1 + sT). Recordando la teoría de transformaciones complejas, esta transformación equivale a una traslación, una inversión y una simetría, que da lugar a un círculo. Al semieje positivo le corresponde el semicírculo negativo dibujado en la figura 20-7. Al semieje negativo le correspondería el semicírculo positivo (trazo interrumpido). El punto $\omega = 0$ tiene módulo K, real, el punto $\omega \to \infty$ es el origen y el punto $\omega = 1/T$ es el complejo K/(1+j), dibujado en la figura (argumento $-\pi/4$).

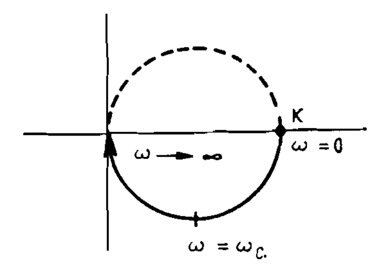


Fig. 20-7.

20.8.- Sistemas de segundo orden.

Son los que están caracterizados por una ecuación diferencial de segundo orden, del tipo:

$$a_0 \frac{d^2y}{dt^2} + a_1 \frac{dy}{dt} + y = Kx$$

normalizado respecto al origen. La función de transferencia será:

$$F(s) = \frac{K}{a_0 s^2 + a_1 s + 1}$$

La respuesta de frecuencia del sistema será:

$$F(j\omega) = \frac{K}{a_0 (j\omega)^2 + a_1 j\omega + 1} = \frac{K}{1 - a_0 \omega^2 + a_1 \omega j}$$

cuyos módulo y argumento son :

$$|F|=\frac{K}{\sqrt{(1-a_0\omega^2)^2+(a_1\omega)^2}}$$

$$\varphi=-\text{arc } tg\frac{a_1\omega}{1-a_0\omega^2}$$

La respuesta de frecuencia es la misma de un filtro RLC en el que a_0 = LC, a_1 = CR. Por consiguiente podemos aplicar todo lo ya estudiado en un capítulo anterior sobre el filtro RLC, en particular la frecuencia o pulsación de resonancia y el factor de calidad Q que define la agudez de la resonancia. Los parámetros del filtro RLC en función de los coeficientes de la ecuación de segundo orden serán:

$$\omega_r=\frac{1}{\sqrt{a_0}} \quad Q=\sqrt{\frac{L}{C}}\frac{1}{R}=\sqrt{\frac{LC}{C^2R^2}}=\frac{\sqrt{a_0}}{a_1}$$

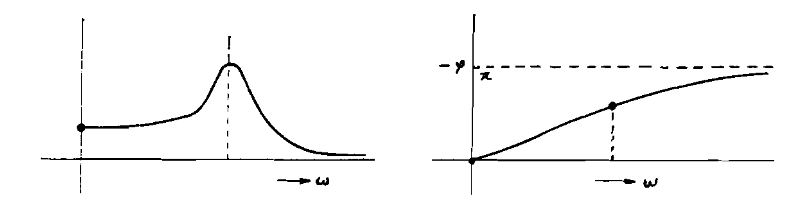


Fig. 20-8.

La representación gráfica es la ya conocida (figura 20-8). El máximo exacto de |F| puede obtenerse derivando el radicando e igualando a cero, con lo que se obtiene:

$$\omega_r=\sqrt{\frac{1}{a_0}\left(1-\frac{a_1^2}{2a_0}\right)}=\sqrt{\frac{1}{a_0}\left(1-\frac{1}{2Q^2}\right)}$$

En el caso de que $a_1^2 << a_0$ (que es el que estudiábamos como Q alta), la pulsación y la frecuencia de resonancia tienden a los valores:

$$\omega_r=\sqrt{\frac{1}{a_0}} \quad f_r=\frac{1}{2\pi}\sqrt{\frac{1}{a_0}}$$

El **diagrama logarítmico de Bode** es el representado en la

figura 20-9 (en amplitud). A frecuencias bajas, la curva $F(j\omega)$ tiende a K, es decir, a log K. A frecuencias altas $F(j.\omega)$ se aproxima a $- K / a_0\omega^2 = - K \omega_r^2 / \omega^2$, o sea $\log |F| = \log K - 2\log(\omega/\omega_r)$ la asíntota es una recta de pendiente -2 que empieza en la frecuencia de resonancia a la altura de K. La aproximación es el diagrama asintótico de Bode, formado por una recta a la altura **log K** y otra que baja a pendiente -2 a partir de log ω_r. La pendiente -2 puede decirse que equivale a que la amplitud "se divide por 4 cuando la frecuencia se duplica", o sea que baja a **"12 dB por octava"**.

El diagrama de Nyquist es el correspondiente a los valores de la figura 20-8 en polares, y tiene la forma que se ve en la figura 20-10. La resonancia aparece en forma de bulto hacia abajo más acusado cuanto mayor sea la relación la "Q" del circuito. El diagrama es tangente al eje real negativo, porque la fase tiende a $-\pi$ para $\omega \to \infty$.

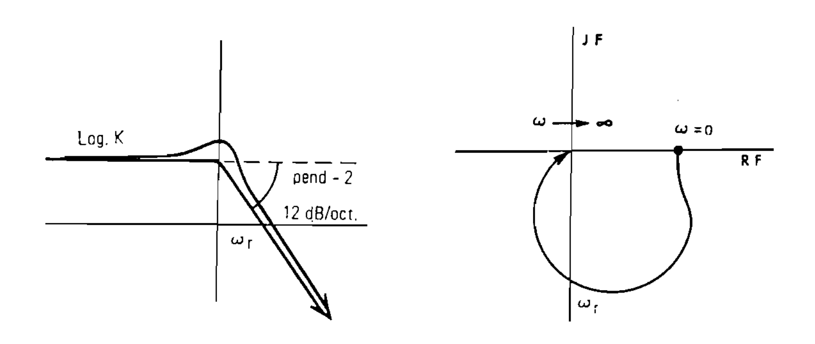


Fig. 20-9. Fig. 20.10.

La **respuesta al escalón unidad** es la transitoria de segundo orden, conocida de la teoría de ecuaciones diferenciales. Corresponde a la transformada de Laplace

$$Y(s) = \frac{K}{s} \frac{1}{a_0 s^2 + a_1 s + 1}$$

La forma depende del signo del discriminante de la ecuación característica. Si el discriminante es negativo, las raíces de la ecuación característica son imaginarias y la solución es oscilatoria, como se ha representado en la figura 20-11. El caso **oscilatorio**, es decir, cuando $a_1^2 - 4 a_0 < 0$ correspondía en los filtros RLC a $1/4Q^2 < 1$, es decir, $Q > 1/2$. La atenuación y la pulsación de los transitorios corresponden a las partes real e imaginaria respectivamente de las raíces de la ecuación característica. Por consiguiente, la atenuación de los transitorios será $-a_1 / 2a_0$ y su pulsación será:

$$\omega_{trans.}=\sqrt{\frac{1}{a_0}-\left(\frac{a_1}{2a_0}\right)^2}=\sqrt{\frac{1}{a_0}\left(1-\frac{a_1^2}{4a_0}\right)}$$

que tiende a ser la de resonancia si Q es alta.

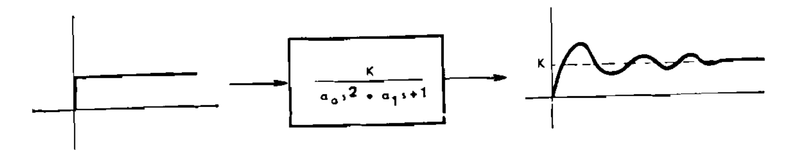


Fig. 20-11.

Para los valores de Q comprendidos entre 1/2 y 1/√2 hay un máximo de la respuesta, aunque ya no es oscilatoria (figura 20-12), y **para valores inferiores de Q la curva se va pareciendo más y más a la respuesta de primer orden.**

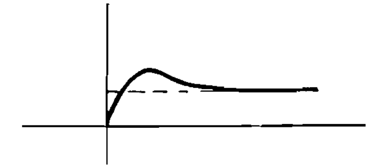


Fig. 20-12.

20.9.- Sistemas de orden superior.

Algunas de las propiedades anteriores pueden generalizarse a sistemas de orden superior. Si la función F(s) es del tipo

$$F(s)=\frac{K}{a_0 s^n+a_1^{n-1}+\ldots+1}$$

la respuesta de frecuencia será:

$$F(j\omega)=\frac{K}{a_0(j\omega)^n+a_1(j\omega)^{n-1}+\ldots+1}$$

y la representación gráfica comenzará (ω = 0) en el punto K para terminar cuando $\omega \to \infty$ tangente al eje que corresponde retrasando n cuadrantes a partir del semieje real positivo (ya que F(jω) será un ∞ equivalente a K/ $a_0\omega^n j^n$.

Los subsistemas estudiados hasta ahora obedecían a una ecuación diferencial en la que el segundo miembro era Kx, con **escala** normalizada al valor K. Estos sistemas se llaman **regulares**, y son los más frecuentes. Pero existen otros sistemas en los que el segundo miembro de la ecuación diferencial es función de x y sus derivadas. Si se llama **n** a la **diferencia de orden** de los dos miembros y se conserva el factor de escala K, como relación de los coeficientes de x e y, se pueden aplicar las dos propiedades anteriores (principio y final del diagrama de Nyquist).

En general, la función racional F(s) puede descomponerse en fracciones simples que corresponden a subsistemas de primero y segundo orden, según las raíces del denominador. Es decir, que aparecerán tantas resonancias como raíces dobles imaginarias tenga el denominador. Esto es más cierto cuanto más separadas estén esas raíces, ya que, en caso contrario, estos efectos se enmascaran unos a otros.

En todos los casos estudiados hasta ahora, hemos supuesto que el coeficiente de y (que hemos normalizado a 1) no es cero, lo que algunos libros llaman **"tipo 0"** o **"sin ninguna integración"**. Sin embargo, el comportamiento a bajas frecuencias varia mucho cuando hay integraciones de tipo 1 (coeficiente de y nulo) o de tipo k cuando el último coeficiente no nulo es el de la derivada k-sima.

Por ejemplo si es de orden 1, $a_n = 0$, $a_{n-1} = 1$ (normalizado), a bajas frecuencias $F(j\omega)$ es un ∞ equivalente a $F(j\omega) = K/a_{n-1}j\omega$ y el diagrama de Nyquist comienza con valor ∞ y argumento $-\pi/2$. En el caso de tipo k (**k integraciones**), $F(j\omega)$ es, a bajas frecuencias, equivalente a $K / a_{n-k}\omega^k j^k$ y el comienzo del diagrama de Nyquist ocurre con módulo ∞ y argumento retrasado k cuadrantes (en sentido de las agujas del reloj) desde el eje real positivo.

20.10. -Retardo puro.

El retardo puro τ es un sistema especial (figura 20-13) cuya función de transferencia es $F(s) = e^{''}$ según el teorema del retardo. Las funciones sinusoidales sufren también un retardo τ, es decir un desfase $\omega\tau$ en radianes. Por consiguiente su transformada compleja se obtiene multiplicando por $e^{-j\omega\tau}$ que es un complejo de módulo 1 y argumento $-\omega\tau$. Queda así justificado que, también para este tipo de funciones, se obtiene la respuesta de frecuencia sin más que sustituir s por $j\omega$.

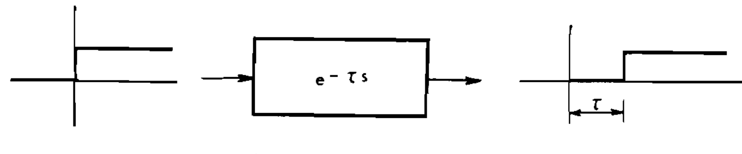


Fig.20-13.

El diagrama de Bode en amplitud es el eje horizontal, 0 dB, ya que la amplitud es siempre 1. El diagrama de fase sería una

recta de pendiente $-\tau$.

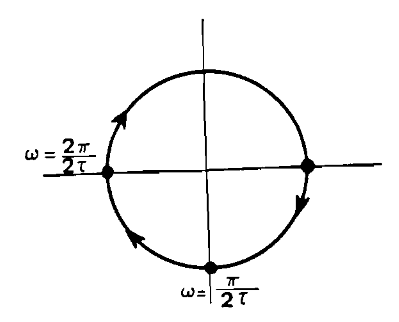


Fig. 20-14.

El diagrama de Nyquist (figura 20-14) es un círculo de radio unidad. El desfase es creciente (sentido de las agujas del reloj) y proporcional a ω. Cortará a los ejes en los valores de ω que hacen $\omega = n\tau / 2$, o sea, $\omega_n = n\pi / 2\tau$. Dará infinitas vueltas.

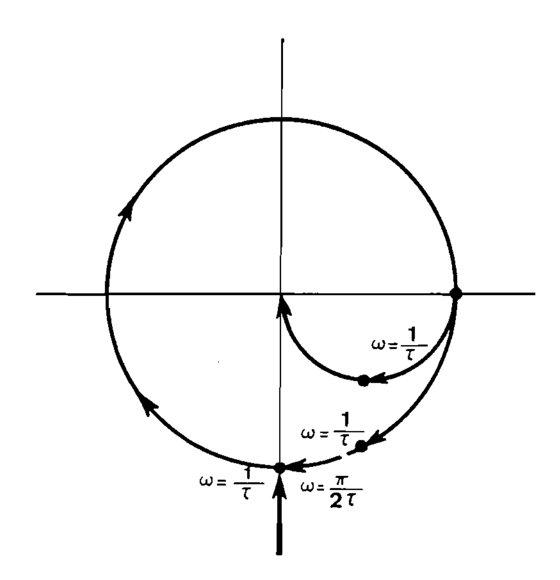


Fig. 20-15.

Es interesante dibujar sobre el mismo papel los diagramas de Nyquist del retardo puro τ, integración de primer orden de constante de tiempo τ e integración pura a la que tendería este último si τ es suficientemente grande (Fig. 20-15). Es decir, los diagramas:

a) Retardo τ. $F = e^{-s\tau}$.

b) Retardo de la ecuación $\tau y' + y = x$ o sea $F = 1 / (1+\tau s)$.

c) Retardo de la ecuación $\tau y' = x$ o sea $F = 1/\tau s$, $F(j\omega) = 1 / j\omega\tau$ que es una "integración pura", cuyo diagrama de Nyquist es el semieje imaginario negativo.

El diagrama de primer orden se parece al de retardo puro a frecuencias bajas (empieza en el mismo punto, 1, con la misma dirección) y acaba pareciéndose, a frecuencias altas, a la integración pura (acaba en el origen tangente al eje imaginario negativo).

20.11.- Función de transferencia de los cuadripolos.

La función de transferencia es una **simplificación** de la matriz del cuadripolo en que se ha supuesto la impedancia de salida cero, quedando sólo la **amplificación**. Es el parámetro A(s), relación salida a entrada, cuando $Z_L = \infty$. Valen por consiguiente todas las respuestas de frecuencia que hemos estudiado que, en realidad, son los diagramas de Bode. En particular:

a) El sistema RC es un sistema de primer orden, de función de transferencia $F(s) = 1/(1 + sRC)$

b) El filtro RLC, pasabajos, es un sistema de segundo orden, cuya función de transferencia y características ya hemos estudiado.

c) El divisor de tensión R_1R_2 tiene función de transferencia constante, $F(s) = R_2 / (R_1 + R_2)$.

d) El circuito de la figura 20-16, llamado de **"avance de fase"** tendrá la función de transferencia:

$$F(s) = \frac{R_2}{R_2 + \dfrac{\dfrac{R_1}{sC}}{R_1 + \dfrac{1}{sC}}} = 1 - \frac{R_1}{R_1 + R_2 + R_1 R_2 sC}$$

que corresponde a la unidad menos una función de primer orden. El diagrama de Nyquist será el de la figura 20-17. La constante de tiempo será:

$$\tau = \frac{R_1 R_2}{R_1 + R_2} C$$

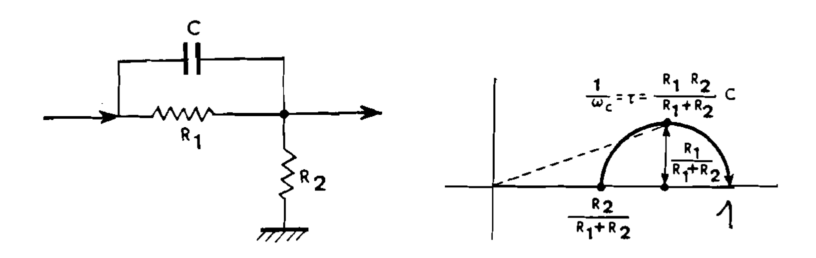


Fig. 20-16. Fig. 20-17.

El desfase a la pulsación $1/\tau$ será (calculado a la vista de la figura 20-17):

$$\varphi=\text{arc } tg\ \frac{\dfrac{2R_1}{R_1+R_2}}{1+\dfrac{R_2}{R_1+R_2}}=\text{arc } tg\ \frac{2R_1}{R_1+2R_2}$$

A cualquier frecuencia hay un **avance de fase positivo**. El avance a la pulsación de corte no es el máximo pero es suficientemente aproximado para las aplicaciones prácticas que estudiaremos más adelante.

20.12.- Función de transferencia de los amplificadores.

El amplificador ideal da una relación salida/entrada constante y, por lo tanto, su función de transferencia será F(s) = K, igual a la amplificación.

En la práctica, el amplificador operacional que se usa en aplicaciones de automática suele ser un amplificador de c.c. con un límite en la curva de frecuencia dado por las capacidades parásitas o bien provocado intencionadamente por un circuito RC para evitar que el comportamiento a altas frecuencias sea errático. Debido a esta técnica, llamada a veces del "circuito dominante", casi todos los amplificadores operacionales se comportan como sistemas de primer orden. La figura 20-18 representa el diagrama de Bode de un amplificador operacional, en que el corte suele fijarse a 10 KHz.(ω_c = 20π o menos.

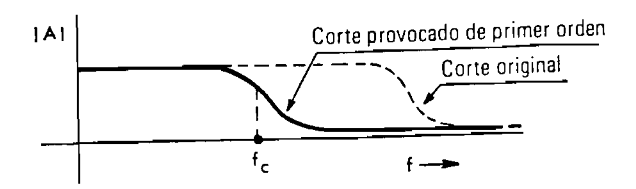


Fig. 20.18.

La función de transferencia será:

$$F(s)=\frac{K}{1+Ts}\quad con\ T=\frac{1}{\omega_c}=\frac{1}{2\pi f_c}$$

La definición dinámica del amplificador se hace con sólo dos parámetros, amplificación y frecuencia de corte.

20.13.- Amplificadores de señal alterna.

En algunos sistemas dé control se utiliza la señal alterna de frecuencia de la red como variable; la variable es en ese caso el valor de la **amplitud** o el valor eficaz de la señal y no su valor instantáneo. A este tipo de sistemas se les llama de "corriente alterna" y son bastante cómodos cuando el actuador es un motor eléctrico de corriente alterna.

Los amplificadores de "corriente alterna" deben amplificar una señal de frecuencia constante, la de la alimentación (generalmente 50 Hz. en red normal o 400 Hz. en aplicaciones aeronáuticas) de forma fiel a la fase, ya que la fase de la señal equivale al "signo" de la corriente continua. Cuando la señal está en fase con la alimentación, diremos que es positiva y cuando está a 180°, diremos que es negativa. Esto es particularmente importante porque define el sentido de giro del motor actuador, en su caso.

Los amplificadores de "alterna" se dividen en dos clases:

a) Amplificadores con entrada continua y salida alterna.

b) Amplificadores con entrada y salida alternas.

En el primer caso, la parte "alterna" de la regulación es la salida del amplificador y el motor, siendo el resto de continua.

Todavía hay que observar otra peculiaridad de ambos tipos de amplificadores, y es que, frecuentemente, la alimentación de la última etapa se hace también alterna, ahorrándose así el

filtrado de la alimentación. Este tipo de amplificador se llama "servoamplificador" y sólo puede usarse con su motor correspondiente.

Aunque hay una gran variedad de circuitos combinando las características mencionadas, veamos uno de cada tipo:

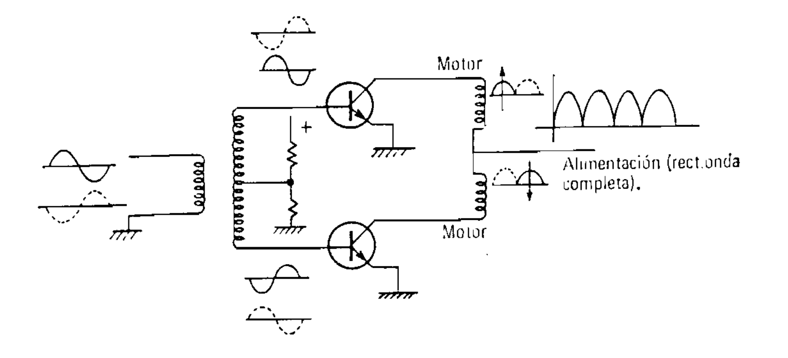


Fig. 20-19.

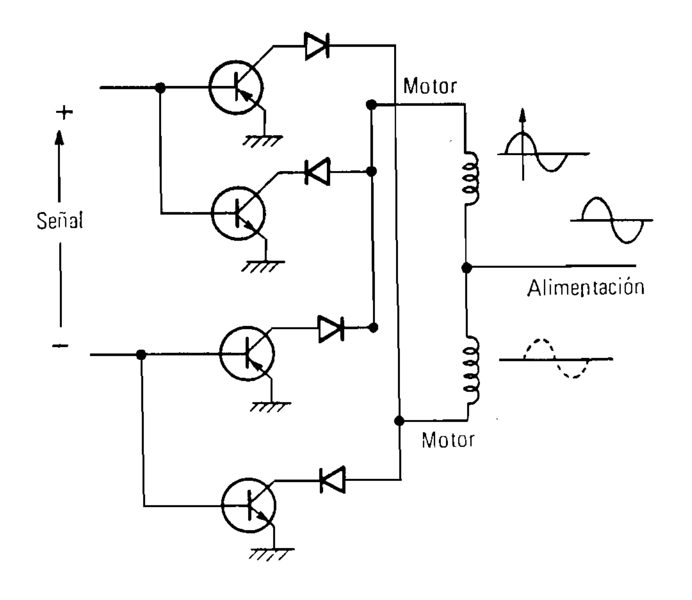


Fig. 20-20.

La figura 20-19 es un "servoamplificador de corriente alterna", es decir, de entrada alterna y salida para motor de alterna. El motor debe tener el devanado cuyo sentido determine

el de giro dividido en dos partes, con una toma central. En la primera parte del ciclo de alimentación conduce el semidevanado superior (suponiendo la señal de entrada como la representada en línea continua) y en la segunda el inferior. La alimentación se hace en alterna rectificada. El resultado suma de ambos semidevanados es como si la señal estuviese en fase con la señal de entrada. Si la fase de la señal es la indicada de trazos, la "salida" está en contrafase.

La figura 20-20 es un servoamplificador de corriente continua, también llamado "amplificador de fase". El amplificador de fase es un circuito diferencial que transforma la señal continua en alterna cuya fase respecto a la alimentación corresponde al signo de la señal. En efecto (figura 20-20), si la señal es positiva con el signo indicado en la figura, la bobina superior conduciría y la bobina inferior no.

Si la polarización de entrada fuese la inversa de la indicada, el efecto seria que durante el primer semiciclo conduciría la bobina inferior a través del npn y durante el segundo a través del pnp. El efecto sería que la señal de salida seria de fase inversa al caso anterior.

La amplitud de la señal depende de la amplificación de los transistores de control; estos amplificadores no suelen ser muy lineales pero sí lo suficiente para las aplicaciones de control en que no se exige más que un valor aproximado de la amplificación.

En cuanto a la respuesta de frecuencia de estos amplificadores, hay que hacer notar que sólo se pueden usar a frecuencias de señal más bajas que la de alimentación. No hay que olvidar que cuando hablemos de respuesta de "frecuencia" nos referimos a la frecuencia con que varía la amplitud de la señal alterna, independiente de la frecuencia de la señal como tal tensión alterna. Así, en los amplificadores de 50 Hz., la máxima frecuencia a que se puede utilizar será de unos 10 c.p.s., es decir la máxima "velocidad" a la que se puede cambiar la señal de 50 Hz. es de 10 c.p.s.(10 Hz.).

Si se intenta usar el amplificador a frecuencias próximas a la de la red, aparecen fenómenos que dependen de cada tipo de amplificador; en general se prefiere limitar la respuesta de frecuencia en otros elementos de la regulación. Por ejemplo, el servoamplificador, con el motor incluido, puede tener una respuesta de frecuencia (tomando como variable de salida la velocidad del motor) de primer orden, como la de la figura 20-18, con un corte a frecuencias del orden 10-20 c.p.s.

20.14.- Amplificadores magnéticos.

El amplificador magnético o de saturación es otro tipo de amplificador basado en el principio de la figura 20-21. El núcleo magnético representado constituye un transformador con un primario (alimentación), un secundario (salida) y dos devanados especiales, uno de control y otro de polarización.

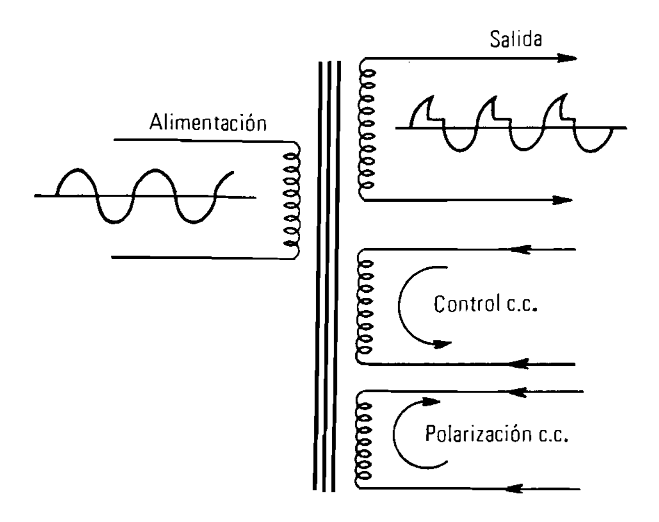


Fig.20-21.

El devanado de control es la entrada del sistema en c.c. El devanado de salida suministra una tensión alterna cuando el núcleo está funcionando como transformador normal, fuera de la saturación. Cuando la corriente de control llega a un valor tal que el núcleo se satura, la salida desaparece porque, al alcanzar el campo magnético el valor de saturación, el flujo deja de variar y desaparece la tensión inducida. Con ello la forma de onda se distorsiona como se indica en la figura 20-21, siendo el punto de "ruptura" de la onda más próximo al cero cuanto más corriente de control se haga pasar. Se obtiene así una dependencia de la salida (valor eficaz) respecto a la corriente de control como la representada en la figura 20-22. Con el devanado auxiliar se puede "polarizar" de forma que se obtenga un punto de trabajo donde la variación del valor eficaz de salida sea máxima al variar la corriente de control.

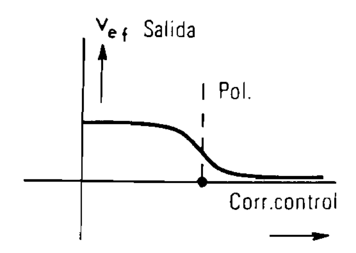


Fig. 20-22.

La señal de salida no es utilizable directamente debido a su gran distorsión, pero puede filtrarse o rectificarse antes de usarla.

Igual que en el caso anterior, la respuesta de frecuencia del amplificador se suele limitar exteriormente, evitando las frecuencias próximas a la de alimentación.

20.15.- Dispositivos de medida electromecánicos.

La mayoría de los dispositivos de medida electromecánicos son dinamómetros que miden la fuerza producida por algún otro efecto físico y la convierten en variable desplazamiento y, de ahí, la transforman en eléctrica. (Ver capítulo de transductores). Veamos algunos ejemplos:

El acelerómetro es una masa suspendida de un muelle que se desplaza por la fuerza de inercia (ver capítulo de transductores). La fuerza de inercia es $M\gamma$; si produce un desplazamiento x, la fuerza será igual a la suma de las fuerzas que actúan sobre M, que serán la de inercia en el movimiento relativo, Mx", la de rozamiento fx' y la del muelle Kx, es decir

$$M\gamma = Kx + fx' + Mx''$$

o sea, en el plano de Laplace

$$\frac{X}{\gamma}=\frac{\frac{M}{K}}{\frac{M}{K}s^2+\frac{f}{K}s+1}$$

que es un sistema de segundo orden en que las constantes valen:

$$a_0=\frac{M}{K}\ ;\ a_1=\frac{f}{K}$$

y la amplificación será M/K. Por consiguiente la pulsación de resonancia propia, la Q del circuito y la atenuación valdrán:

$$\omega_r=\sqrt{\frac{1}{a_0}}=\sqrt{\frac{K}{M}}$$

$$Q=\frac{\sqrt{a_0}}{a_1}=\frac{\sqrt{\frac{M}{K}}}{\frac{f}{K}}=\frac{1}{f}\sqrt{MK}$$

$$\alpha=\frac{\frac{f}{K}}{2\frac{M}{K}}=\frac{f}{2M}$$

Si la salida del dinamómetro se conecta a un potenciómetro

el efecto es un aumento de f (frotamiento en el potenciómetro), conservándose la misma función de transferencia con salida eléctrica.

Los voltímetros, amperímetros y otros instrumentos de medida electromecánicos tienen la misma respuesta de frecuencia de segundo orden, ya que todos ellos son dinamómetros en que se mide la fuerza ejercida por un campo magnético sobre una corriente. Por ello su función de transferencia es la ya citada, y está caracterizada por tres parámetros que pueden ser la escala K, la resonancia ω_r y la atenuación α. En el caso de que los parámetros vengan dados de esa forma, se obtienen a_0 y a_1, resolviendo las ecuaciones aproximadas:

$$\omega_r=\frac{1}{\sqrt{a_0}}\ ;\ \alpha=\frac{a_1}{2a_0}\ \text{de donde}\ a_0=\frac{1}{\omega_r^2}\ ;\ a_1=\frac{2\alpha}{\omega_r^2}$$

Otros transductores tienen también respuesta de segundo orden, como el giróscopo, etc. (se pueden aproximar a segundo orden).

20. 16.- Motores eléctricos de c.c.

Consideremos el motor eléctrico de c.c. de la figura 20-23 en que el inductor está alimentado con corriente constante, es decir, el campo es constante, y el inducido está alimentado con una señal de control.

Sea e la señal de control, i la intensidad del inducido, cuya resistencia y autoinducción llamaremos R y L. Las ecuaciones (aproximadas) del motor expresarán que el par motor T es proporcional a la intensidad i, y que la fuerza contraelectromotriz es proporcional a las revoluciones. Sean éstas $d\theta/dt$ (siendo θ el ángulo girado) y sean K_1 y K_2 las dos constantes de proporcionalidad.

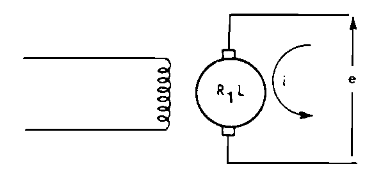


Fig. 20-23.

Las ecuaciones electromecánicas, si llamamos J al momento de inercia y f a la fricción del motor, serán:

$$T=K_1 i=J\frac{d^2\theta}{dt^2}+f\frac{d\theta}{dt}$$
$$e-K_2\frac{d\theta}{dt}=Ri+L\frac{di}{dt}$$

Pasando al plano de Laplace, las ecuaciones serán:

$$T=K_1 i=Js^2\theta+fs\theta$$
$$e=K_2 s\theta+Ri+Lsi$$

donde sustituyendo el valor de i en la última y despejando θ en función de e:

$$\frac{\theta}{e}=\frac{K_1}{LJs^3+(RJ+Lf)\,s^2+(Rf+K_2K_1)\,s}$$

El motor eléctrico es pues un sistema de tercer orden si se toma como variable de salida la posición, con una integración pura, que corresponde simplemente a que la posición es la integral de las revoluciones. Si se toma como variable de salida la velocidad angular, el sistema queda de segundo orden. Si lo identificamos con la función de transferencia de segundo orden podremos calcular el valor de K, amplificación, en que, como vemos en la expresión, se pone de manifiesto en su denominador el efecto desfavorable de la fricción, que necesita un par para vencerla que no se transforma en velocidad:

$$F(s)=\frac{\theta s}{e}=\frac{K}{a_0s^2+a_1s+1}$$
$$K=\frac{K_1}{Rf+K_1K_2}=\frac{1}{\frac{Rf}{K}+K_2}$$

Los demás coeficientes se identifican así:

$$a_0=\frac{LJ}{Rf+K_1K_2}\quad \omega_r=\sqrt{\frac{Rf+K_1K_2}{LJ}}$$
$$a_1=\frac{RJ+Lf}{Rf+K_1K_2}\quad \alpha=-\frac{RJ+Lf}{2LJ}$$

En el caso de que L sea despreciable (como lo es normalmente), el sistema queda de primer orden en la velocidad y de segundo en la posición. En ese caso, tomando como variable

de salida la velocidad:

$$F(s)=\frac{\theta s}{e}=\frac{K_1}{RJs+Rf+K_1K_2}=\frac{K}{\tau s+1}$$

el valor de K es el mismo que sin L despreciable y la constante de tiempo vale:

$$\tau=\frac{RJ}{Rf+K_1K_2}$$

Si el rozamiento es muy bajo, es decir si f= 0, la función de transferencia se reduce a

$$F(s)=\frac{K_1}{RJs+K_1K_2}$$
$$K=\frac{1}{K_2}\ y\ \tau=\frac{RJ}{K_1K_2}$$

y si además R o J son despreciables,

$$F(s)=\frac{\theta s}{e}=\frac{1}{K_2}\quad \frac{\theta}{e}=\frac{1}{K_2 s}$$

En ese caso el motor se convierte en un integrador puro respecto a e, aproximación que, a veces, es suficientemente buena para los motores muy pequeños que posicionan algún elemento del sistema sin aportación apreciable de potencia.

Los motores de c.a. tienen un comportamiento parecido cuando se utilizan como "universales". Es poco corriente la utilización de motores de jaula de ardilla como elementos del servomecanismo, salvo en casos de regulación por conmutación.

CAPITULO 21

ACTUACIONES DE LOS SISTEMAS DE REGULACION AUTOMATICA

21.- Introducción.

En los capítulos anteriores se ha esbozado el problema del control, se ha restringido a su forma elemental y se han tratado las características de los subsistemas.

Ahora vamos a estudiar el comportamiento dinámico de los sistemas en circuito cerrado. Fundamentalmente nos interesa la estabilidad y las actuaciones de los mismos.

Estudiaremos la estabilidad por el método de Nyquist. Es clásico estudiarlo por este método o por el de las raíces, a veces llamado de Evans. No obstante, aquí nos limitaremos al primero, que consideramos muy superior por no necesitar un conocimiento matemático tan completo de los subsistemas. Por el mismo motivo no trataremos tampoco los otros métodos existentes, como el estudio de la ecuación diferencial (discriminante de Raouth), etc.

21.2.- Estabilidad.

La estabilidad de los sistemas lineales puede definirse así: Diremos que un sistema (lineal) es estable si, ante cualquier perturbación transitoria, tiende a una situación de reposo.

Si el sistema es lineal, se podrá tomar como momento inicial el del fin de la perturbación; el comportamiento del sistema estará entonces definido por una ecuación diferencial lineal con valores iniciales no nulos y sin perturbación. La solución estacionaria será la de reposo y quedará sólo la solución transitoria, que debe tender a cero, ya que sólo admite soluciones convergentes, oscilatorias o divergentes. Por

consiguiente, la condición de estabilidad será que la ecuación diferencial que define el comportamiento del sistema solo tenga regímenes transitorios convergentes o, lo que es lo mismo, como se sabe por matemáticas:

La condición necesaria y suficiente para que un sistema sea estable es que la ecuación característica de la ecuación diferencial que debe cumplir su variable de estado tenga todas sus raíces negativas o de parte real negativa. Con ello se garantiza que las soluciones transitorias sean convergentes. Estas raíces, en el caso de las funciones racionales son las del denominador, es decir, los polos de F(s).

21.3.- Función de transferencia en circuito cerrado.

Consideremos primero el caso de la realimentación unitaria (figura 21-1). La función de transferencia F(s) en circuito abierto será:

$$F(s)=\frac{x(s)}{x_0(s)-x(s)}$$

siendo x(s) la transformada de Laplace de la salida del sistema (abierto o cerrado) y x_0(s) la del valor deseado o "de referencia" de x(s). (Se entiende que, siempre que en una fórmula figure la función de transferencia o las variables sean función de "s", estamos en el plano complejo, todas las variables son transformadas de Laplace y el tiempo "t" no debe aparecer como variable).

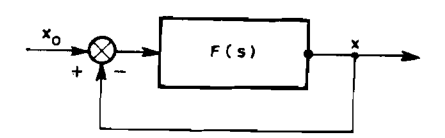


Fig. 21-1.

Se llama **función de transferencia en circuito cerrado, G(s)** al valor

$$G(s)=\frac{x(s)}{x_0(s)}$$

G(s) es la relación entre la salida o variable de estado del sistema cerrado y la entrada del sistema cerrado o valor de referencia. (La salida de los sistemas cerrado y abierto es la misma, x(s), pero no así la entrada, x_0(s) - x(s). Despejando de F(s) se obtiene:

$$F(s)x_0(s)-F(s)x(s)=x(s)\text{, o sea}$$
$$G(s)=\frac{x(s)}{x_0(s)}=\frac{F(s)}{1+F(s)}$$

En esta fórmula ya se pone de manifiesto la naturaleza de la regulación : si **F(s) es muy grande** (módulo grande), G(s) tiende a 1, y **la salida sigue a la entrada,** típico de la **realimentación unitaria.**

Si la **realimentación no es unitaria**, sino que vale ,β(s) podemos escribir (figura 21-2):

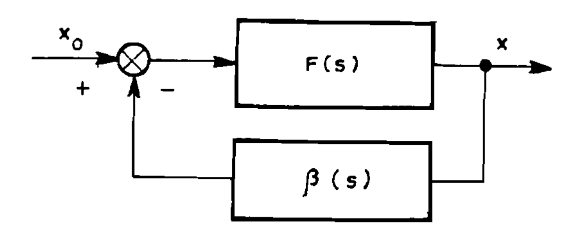


Fig. 21-2.

$$F(s)=\frac{x(s)}{x_0(s)-\beta(s)x(s)}\text{ o sea }F(s)x_0(s)-F(s)\beta(s)x(s)=x(s)$$
$$G(s)=\frac{x(s)}{x_0(s)}=\frac{F(s)\beta(s)}{1+F(s)\beta(s)}\frac{1}{\beta(s)}$$

Ahora, la **función de transferencia en circuito cerrado,** que vale **F(s) β(s)** es distinta de la del sistema F(s). La primera fracción del resultado de G(s) nos dice que el seguimiento se hace como antes, siempre que F(s)β(s) sea grande, pero **la salida tiende a valer** entonces la segunda fracción **1/ β(s). La transformación que hace el sistema es 1/ β(s),** la inversa de la realimentación. La figura 21-3 pone de manifiesto este efecto.

En general **β es la transformación de medida** (que debería ser 1) y no es de suponer que introduzca una inestabilidad, por lo que la estabilidad del sistema será la del sistema simplificado de realimentación unitaria. En cuanto a la precisión y actuaciones habrá que contar con β en todo caso, que se hace así la parte más sensible del sistema, ya que la realimentación no corrige los errores de β.

En adelante estudiaremos sólo sistemas con realimentación unitaria. Los de realimentación no unitaria deben estudiarse como de realimentación unitaria y función de transferencia Fβ, considerando por separado la transformación 1/β.

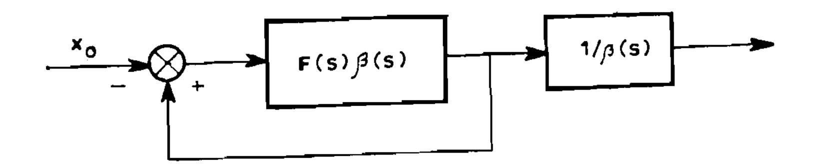


Fig. 21-3.

21.4.-Estabilidad en circuito cerrado.

La función de transferencia en circuito cerrado G(s) debe tener todos sus polos negativos o de parte real negativa. Estos polos son los ceros o raíces de la ecuación característica del denominador 1 + F(s).

Los polos de la función F(s) no lo son de G(s), ya que hacen G(s) → 1 y no producen inestabilidad. Es decir, que puede ocurrir (de hecho es frecuente) que F(sJ sea "inestable" (tenga transitorios divergentes y el sistema G(s) sea estable. Incluso, en el plano de la frecuencia, aquellas frecuencias que hacen F(s)→∞ dan una gran exactitud a la regulación, ya que G(s) tiende a 1.

Los ceros de la función F(s) tampoco lo son de 1 + F(s) y tampoco producen inestabilidad, si bien aquellas frecuencias que anulan F(s) dan una mala respuesta (G → 0) al circuito cerrado.

A continuación vamos a ver el método para el análisis de la estabilidad de G(s), basado en el estudio de los ceros de la función 1 + F(s).

21.5.- Criterio de Nyquist.

En primer lugar enunciaremos un teorema de Cauchy relativo a la función M(s) que justificaremos en el caso de que sea racional:

Si se recorre una vez en el plano s el contorno C de un recinto cerrado, en el que la función M(s) tenga P polos y Z ceros y ningún polo ni cero en el contorno, en el sentido negativo-trigonométrico (de las agujas del reloj), el argumento de la función M(s) aumenta $2\pi(P-Z)$.

En efecto, si la función M(s) es

$$M(s) = \frac{(s-Z_1)(s-Z_2)\dots(s-Z_Z)}{s-P_1)(s-P_2)\dots(s-P_P)}$$

su argumento es la suma de los argumentos de los términos del numerador menos los del denominador. Ahora bien, cada término (s-Z_n) es un complejo que podemos representar en el plano s por el

vector que va de Z_n a s. Si Z_n está dentro de C (figura 21-4), al rodear s el contorno, el vector $s-Z_n$ disminuye su argumento en 2π (si rodeamos en sentido "negativo", es decir en el de las agujas del reloj). Si el cero está fuera como Z_k de la figura, al rodear s el contorno $s-Z_k$ vuelve al mismo argumento sin cambiarlo. Por consiguiente aquellos complejos del tipo $s-Z_n$ que corresponden a ceros internos contribuyen disminuyendo 2π la suma de los argumentos que da el de M. Lo mismo ocurre con los polos, pero como están en el denominador, al disminuir el argumento de los polos internos en 2π aumenta 2π el de M. **Por consiguiente el cambio total del argumento de M será $2\pi(P-Z)$.**

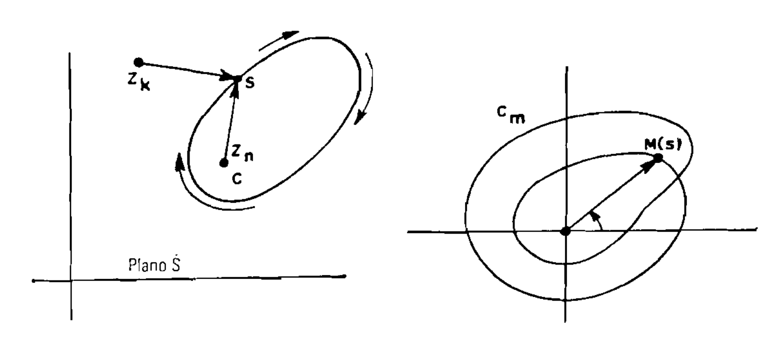


Fig. 21-4. Fig. 21-5.

Si ahora consideramos M(s) como una transformación compleja del plano s al plano M, a un contorno C_s en el plano s le corresponderá otro contorno C_m en el plano M. Al recorrer s el contorno C, M(s) recorrerá el contorno Cm (figura 21-5). Que el argumento de M "cambia" $2\pi(P-Z)$ quiere decir que el punto M da P-Z vueltas en sentido positivo al origen.

Definamos ahora el **"contorno de Nyquist"** (figura 21-6) formado por el eje imaginario (o "$j\omega$") y un semicírculo de radio infinito que se cierra por la derecha. Dentro de este contorno están comprendidos todos los valores de s reales positivos o de parte real positiva. Si consideramos F(s) (función de transferencia de un sistema en circuito abierto) como transformación compleja, el transformado del semieje imaginario positivo es el diagrama de Nyquist ($F(j\omega)$). El transformado del semieje negativo es el complejo conjugado del diagrama de Nyquist, es decir, el diagrama simétrico respecto al eje real (sin embargo, el sentido de recorrido no es el simétrico, ya que en el contorno, ver figura 21-6, el sentido de la flecha es antisimétrico y así lo es en el plano F, cerrando el contorno). El transformado del semicírculo de radio infinito será el valor $F(\infty)$ que, en el caso representado, se ha supuesto cero.

Apliquemos ahora el teorema anterior a la función 1 + F(s) (cuyos ceros nos interesan desde el punto de vista de la estabilidad). La transformación compleja 1 + F(s) dará lugar en el plano (1 + F) al diagrama de Nyquist más el simétrico, desplazado + 1 (hacia la derecha). En la figura 21-7 está

representado el diagrama desplazado para un sistema de segundo orden, como el de la figura 21-6. Supondremos que no hay polos ni ceros de 1 + F en el contorno de Nyquist. El teorema dice que al recorrer el contorno de Nyquist en el sentido de la figura (ω creciente), el número de vueltas del diagrama 1 + F (Nyquist desplazado) alrededor del origen será igual al número de polos menos el número de ceros de 1 + F dentro del contorno, es decir, positivos o de parte real positiva.

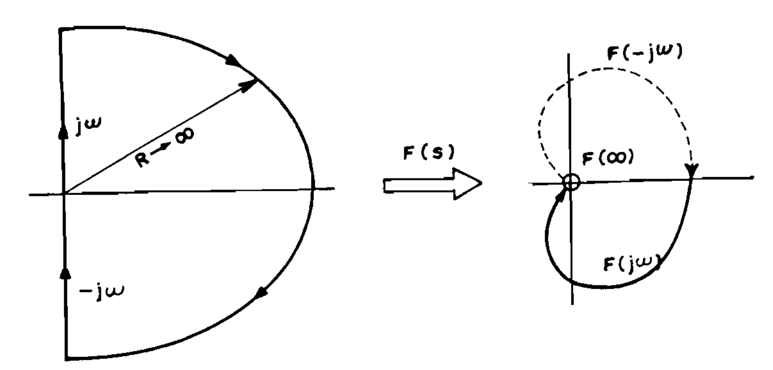


Fig. 21-6.

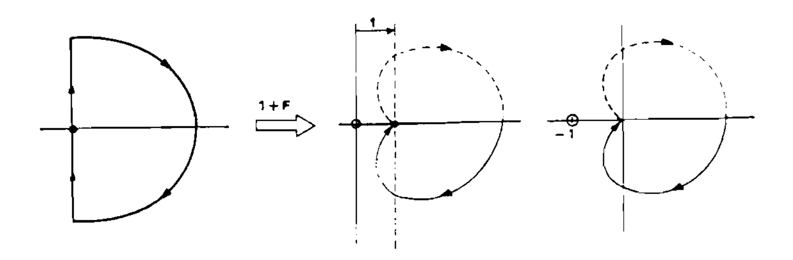


Fig. 21-7.

Ahora bien, el número de vueltas alrededor del origen de 1 + F es lo mismo que el número de vueltas de F(s) alrededor del punto -1 (ver figura 21-7). Este número de vueltas es P-Z, luego para que Z = 0 (caso estable), debe ser el número de vueltas igual a P. De ahí el enunciado del criterio de Nyquist.

La condición necesaria y suficiente para que un sistema sea estable es que su diagrama de Nyquist completo (con el conjugado) de, en sentido de ω creciente, un número de vueltas alrededor del punto critico, -1 igual al número de polos inestables (positivos o de parte real positiva) de la función de transferencia en circuito abierto. Las vueltas se contarán en sentido trigonométrico positivo.

Con este criterio se ve inmediatamente la estabilidad **en circuito cerrado** del examen de la respuesta **en circuito abierto.**

21.6.- Polos de la función de transferencia.

Hemos supuesto que 1 + F no tenía ceros ni polos en el contorno de Nyquist. Veamos qué ocurre si los tiene.

Si 1 + F se hace cero en algún punto del contorno, el sistema es ya inestable, porque G se hace ∞ a esa frecuencia. Por consiguiente se hace innecesaria la aplicación del criterio. Corresponde al caso límite de que el diagrama de Nyquist **pase por el punto crítico** y es, por tanto, **inestable**, notándose con la simple inspección del diagrama.

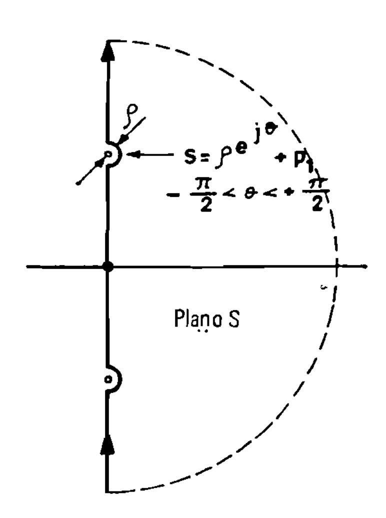


Fig. 21-8.

Si 1 + F se hace ∞ en algún punto del contorno, es decir si F se hace ∞ (ya que 1 + F tiene los mismos polos que F), no se puede aplicar el criterio. Para eludir esta dificultad, que se presenta frecuentemente, sobre todo para s = 0, se usa la teoría de residuos. En lugar de recorrer el contorno de Nyquist, se eluden los polos con **pequeños semicírculos de radio ρ**, dejando fuera los polos de F (figura 21-8). Por ejemplo el polo p_1 de la figura 21-8 se elude con un semicírculo de radio ρ. El valor de s en ese semicírculo será:

$$s=p_1+\varrho e^{j\theta}$$

donde ρ es muy pequeño y θ se hace variar de $-\pi/2$ a $+\pi/2$. Naturalmente, el valor de $F(p_1)$ es ι; al sustituir s en F(s) quedará un resto o infinito equivalente función de ρ y θ, que nos dé la forma del diagrama de Nyquist en las proximidades. Veamos ahora ejemplos típicos de aplicación.

Tomemos como ejemplo la función:

$$F(s) = \frac{1}{s^2+a^2}$$

en circuito abierto, que corresponde a la respuesta de frecuencia

$$F(j\omega) = \frac{1}{s^2-\omega^2}$$

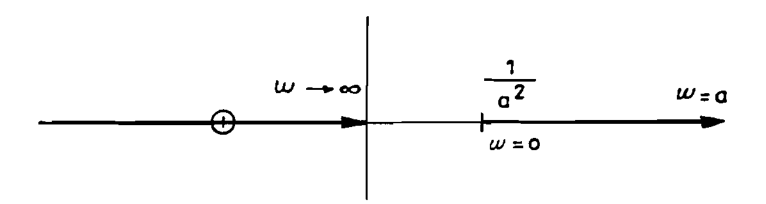


Fig.21-9.

La función de transferencia F(s) tiene los polos s =±ja en el contorno de Nyquist. Estos polos son dos puntos que hacen F→∞ en el diagrama de Nyquist, que será el de la figura 21-9. Sin embargo, el diagrama pasa necesariamente por -1 (cuando $\omega^2 = a^2 + 1$), luego ya es inestable.

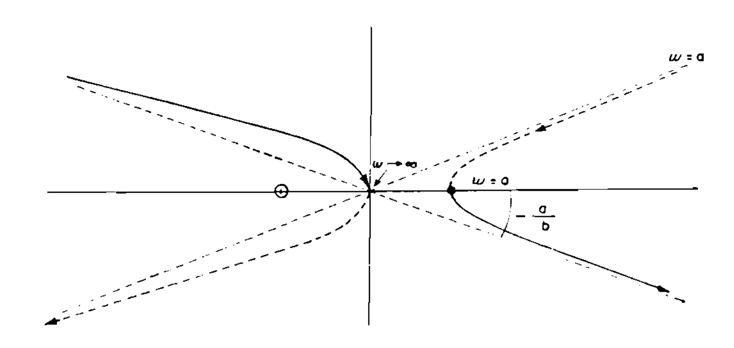


Fig. 21-10.

Tomemos ahora la función:

$$F(s) = \frac{1}{(s^2+a^2)\ (s+b)}$$

que tiene un polo s =-b "estable", que no cuenta para la

aplicación del criterio, y dos polos s=±ja en el contorno de Nyquist. La respuesta de frecuencia será:

$$F(j\omega)=\frac{1}{(a^2-\omega^2)\,(j\omega+b)}$$

y su representación, diagrama de Nyquist, es el de la figura 21-10. En este caso, los dos polos producen dos asíntotas de pendiente ±a/b y no es posible contabilizar las vueltas que da el diagrama alrededor del punto crítico. Recorremos ahora el contorno de Nyquist como se indica en la figura 21-8, dejando fuera los polos s = ± aj. Sustituyendo el valor s = ρ $e^{j\theta}$ + aj en la función F(s), se obtiene:

$$F(s)=\frac{1}{(\varrho^2e^{2j\theta}+2\varrho e^{j\theta}aj-a^2+a^2)\,(\varrho e^{j\theta}+aj+b)}$$

en que despreciando infinitésimos de segundo orden frente al primero y de primer orden frente a aj + b, se obtiene:

$$F(s)\to\frac{1}{2a\varrho je^{j\theta}(b+aj)}=\frac{-1}{2a\varrho\,(b+aj)}\,e^{-j\theta}$$

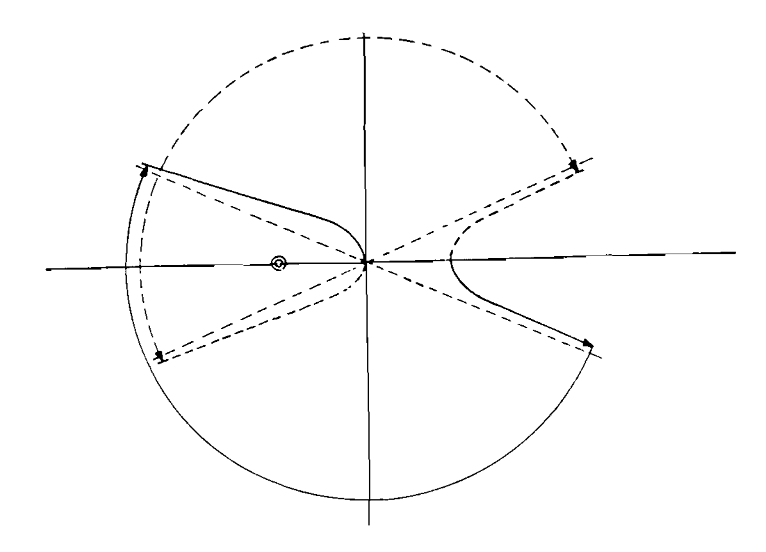


Fig. 21-11.

F(s) será (en las proximidades de s = aj un semicírculo de radio ∞, ya que su módulo será constante, y su argumento cambia

π cuando θ varíe de $-\pi/2$ a $+\pi/2$. Como $-\theta$ varía entre $+\pi/2$ y $-\pi/2$, el semicírculo de radio ∞ gira en el sentido de las agujas del reloj (negativo) lo que cierra la figura como se indica en la figura 21-11 (ω = aj). Del mismo modo, para s=-aj, se obtiene otro semicírculo con el mismo sentido de giro. El número de vueltas alrededor del punto crítico es ahora 2, y el sistema (que no tiene, en circuito abierto, ningún polo dentro del contorno de Nyquist) es inestable.

21.7.- Observaciones sobre el criterio de Nyquist.

Los sistemas de primero y segundo orden que no tienen inestabilidad en el circuito abierto, son estables siempre en circuito cerrado, ya que sus diagramas de Nyquist nunca encierran al punto critico. A partir del tercer orden (y siempre en el caso de que no haya polos inestables en circuito abierto) hay que tener cuidado con el modo de aproximarse al origen. En los sistemas "regulares", en que la forma es del tipo de la figura 21-12, sin asíntotas y decreciente en su última parte, el criterio se puede aplicar en forma simplificada diciendo que el diagrama no debe rodear al punto critico. Esta condición se puede expresar también diciendo que el diagrama debe pasar a la derecha del punto crítico, recorrido en sentido de las ω crecientes. Si esto lo expresamos en otra forma, podemos decir que la condición de estabilidad es que el desfase sea menor de π para amplificación unidad (punto A de la figura 21-13) o que la "amplificación" (módulo de F(jω) debe ser menor que 1 cuando el desfase sea π.

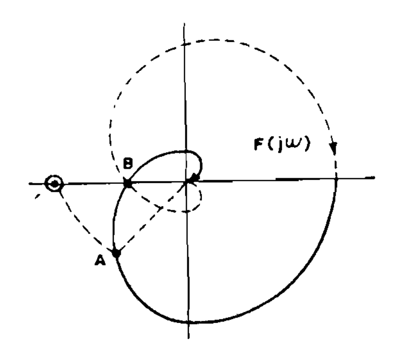


Fig. 21-12.

Así se puede leer también este criterio en el diagrama de Bode (figura 21-13). No olvidar que esta aplicación no es rigurosa y puede haber sistemas inestables que cumplan estas condiciones, aunque son raros.

El retardo puro F(s) = $e^{-\tau s}$ no es una función racional, pero cumple el teorema que dio lugar al criterio de Nyquist. Por esto, se puede aplicar también el criterio de Nyquist al caso en que haya un retardo puro.

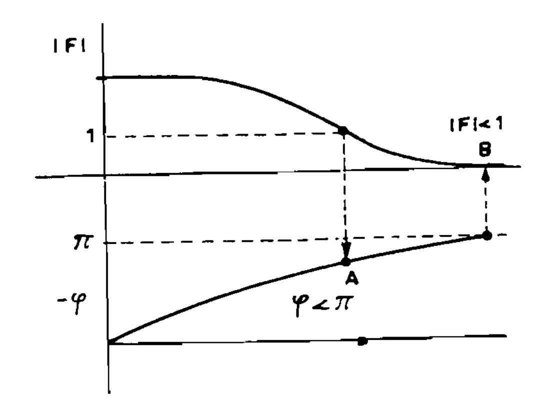


Fig. 21-13.

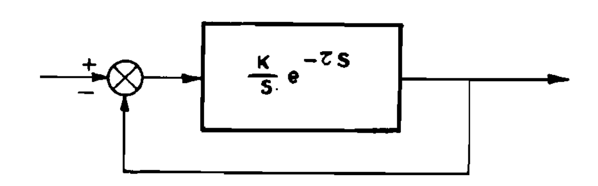


Fig. 21-14.

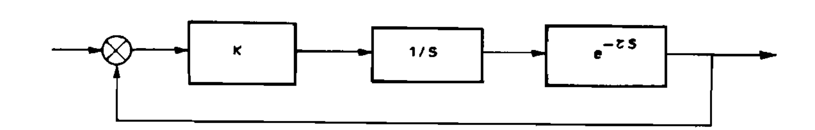


Fig. 21-15.

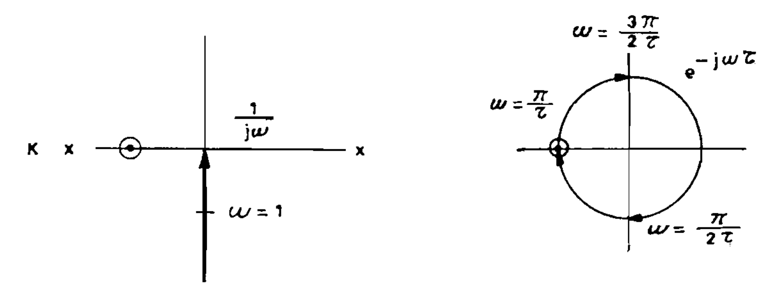


Fig.21-16.

Tomemos como ejemplo el sistema de la figura 21-14, que se podría componer de un amplificador K, de un integrador puro y una línea de retardo τ (figura 21-15). La función de transferencia

y la respuesta de frecuencia serán:

$$F(s) = \frac{K}{s} e^{-\tau s} \quad F(j\omega) = \frac{K}{j\omega} e^{-j\omega\tau}$$

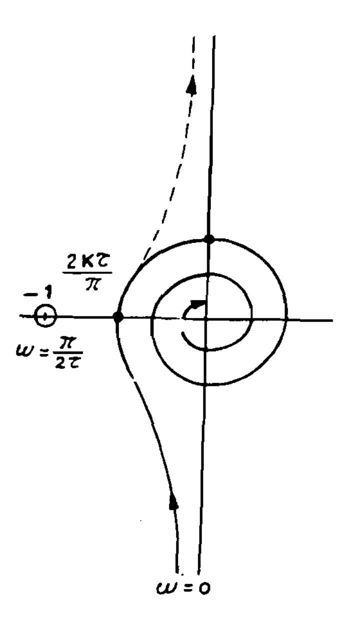


Fig. 21-17.

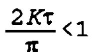
La respuesta será el producto de las representadas en la figura 21-16, dando lugar al diagrama de Nyquist de la figura 21-17. El módulo es K/ω y el argumento decrece indefinidamente, formando una espiral alrededor del origen.

Para aplicar el criterio tenemos que dejar fuera del contorno de Nyquist el punto s = 0 que es el único polo de F(s) (figura 21-18).

Dando a s el valor $\rho e^{j\theta}$, se obtiene:

$$F(s) = \frac{K}{\varrho e^{j\theta\tau}} \rightarrow \frac{K}{\varrho} e^{-j\theta}$$

es decir un semicírculo que se cierra por la derecha (sentido -θ), como se ve en la figura 21-18. La condición de estabilidad será evidentemente:



$$\frac{2K\tau}{\pi}<1$$

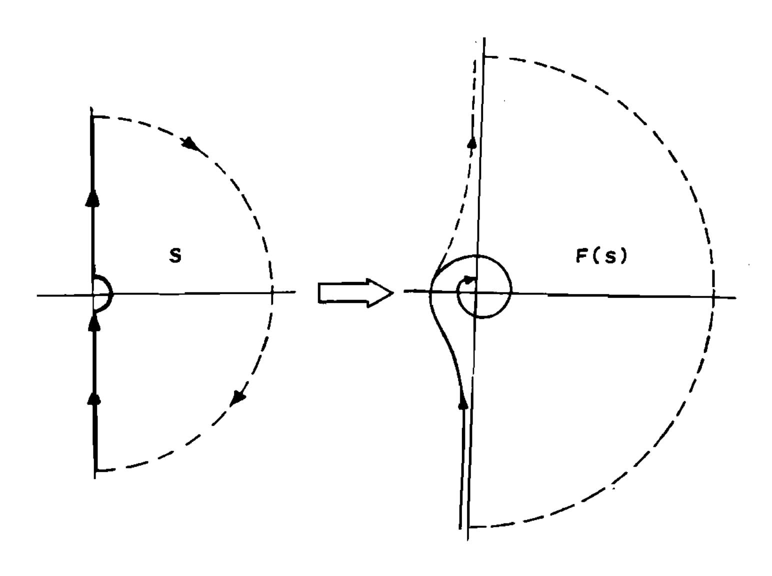


Fig. 21-18.

ya que el diagrama no debe dar ninguna vuelta alrededor del punto crítico (ver también la figura 21-17).

21.8.--Comportamiento en circuito cerrado, ábaco de Hall.

Recordemos que la respuesta de frecuencia en circuito cerrado es:

$$G(j\omega)=\frac{F(j\omega)}{1+F(j\omega)}=\frac{-F(j\omega)}{-F(j\omega)-1}$$

Si observamos el diagrama de Nyquist, notaremos que se puede representar el complejo -F por el vector que va desde cada punto al origen (figura 21-19) y el -F-1 por el vector que une el punto con el punto crítico -1. El complejo G es la relación de estos dos, de donde resulta:

El módulo de la respuesta en circuito cerrado es la relación de distancias a los puntos 0 y -1.

El argumento es el ángulo con el que se ve el segmento -1, 0.

Si representamos sobre el plano del diagrama de Nyquist los círculos lugar geométrico de módulo de G constante (relación de distancias a 0 y -1, constante) y argumento de G constante

(ángulo con el que se ve -1,0 constante) se obtienen así dos familias de círculos ortogonales en los que se **leen** los valores de G y arg G y se llaman **ábaco de Hall.**

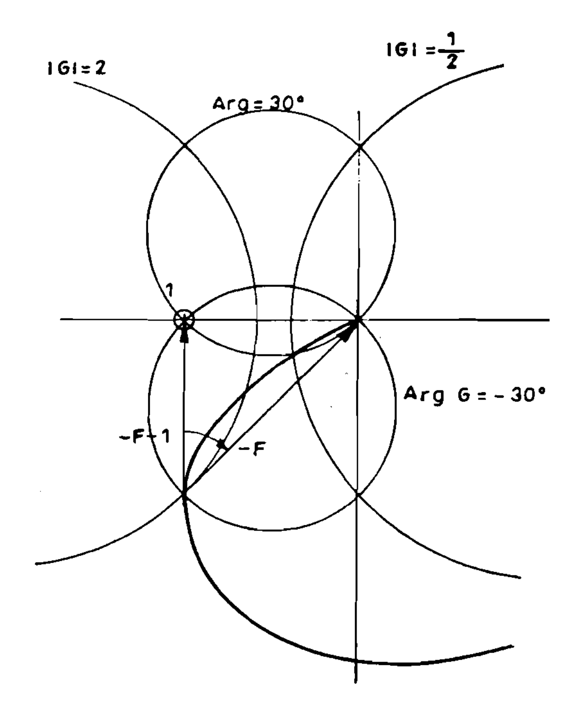


Fig. 21-19.

La función de transferencia en circuito cerrado es, por definición $G(s) = x(s)/x_0(s)$, es decir, la relación entre el valor obtenido y el valor deseado de la variable de estado. Su valor óptimo es la unidad. Como, por otra parte

$$G(s) = \frac{F(s)}{1+F(s)} \quad G(j\omega) = \frac{F(j\omega)}{1+F(j\omega)}$$

deduciremos que $G(j\omega) \to 1$ si $F(j\omega) \to \infty$. En general el sistema regula haciendo F grande para que G se aproxime a 1. Cuanto más grande se pueda hacer F, dentro de la estabilidad, mejor.

En el plano de la frecuencia se puede escribir el error relativo de la variable de estado (o, en función de $F(j\omega)$:

$$\frac{x_0-x}{x}=\frac{1}{G(j\omega)}-1$$

$$\frac{x_0-x}{x}=\frac{1+F(j\omega)}{F(j\omega)}-1=\frac{1}{F(j\omega)}$$

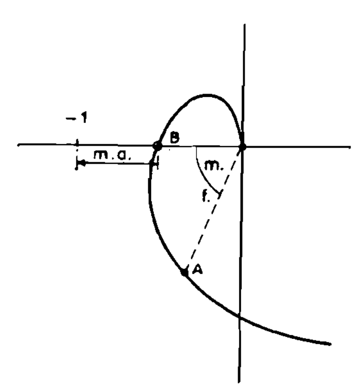


Fig. 21-20.

21.9.- Margen de amplitud y fase, resonancia.

En la mayoría de los sistemas regulados, la inestabilidad se produce cuando el diagrama de Nyquist llega a sobrepasar el punto crítico. La distancia a este punto es una garantía de estabilidad. Así (ver figura 21-20) la distancia en ángulo entre el punto de amplificación unidad y el punto critico se llama **margen de fase.** La distancia del punto de desfase $-\pi$ (B de la figura) al punto critico se llama **margen de amplitud.** En la práctica, el margen de fase debe ser superior a 30° y el de amplitud superior a 0,25.

También es interesante la frecuencia de resonancia que es la que hace que G sea máximo, es decir corresponde al punto de relación de distancias a 0 y -1 máximo. (Suponemos que son sistemas "regulares" en los que estos parámetros son fáciles de identificar).

El punto de resonancia exacta ω_1 sería aquél en que el diagrama es tangente a un círculo del ábaco de Hall (ver figura 21-21), ya que daría la máxima relación de distancias a -1 y 0. En la práctica, este punto es muy aproximado al punto más próximo a -1 (ω_2 del diagrama) e incluso, si el diagrama sube casi perpendicular al eje real, al punto de corte con dicho eje ω_3. Estos tres puntos coinciden prácticamente si el diagrama es del tipo del representado en la figura 21-22, no siendo posible

diferenciarlos. Es muy corriente tomar ω_3 como una buena aproximación.

En la figura 21-21 se ha representado también la resonancia de F, ω_4 para indicar que no está relacionada con la de G; normalmente, la resonancia en circuito cerrado es mucho más alta que la de circuito abierto.

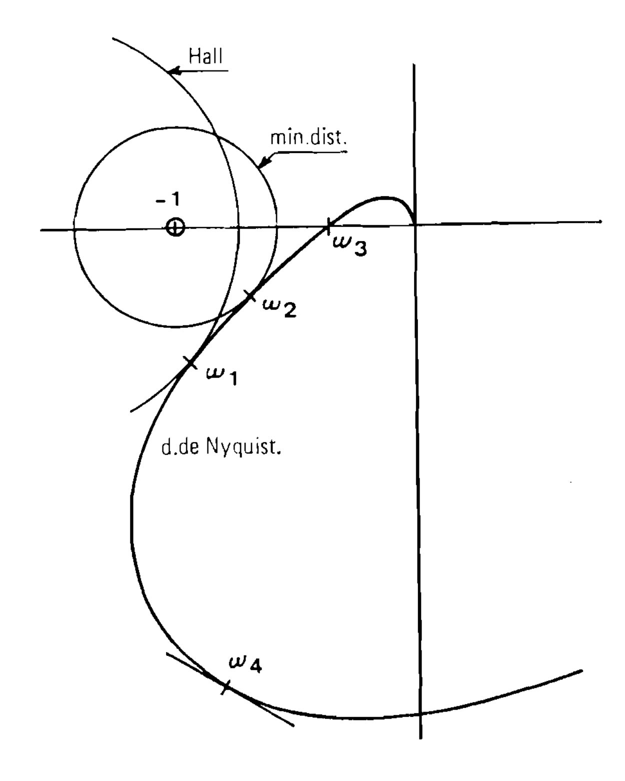


Fig. 21-21.

21.10.- Comportamiento a bajas frecuencias.

Como ya hemos indicado, la respuesta a cualquier frecuencia tiene por módulo la relación de distancias a los puntos 0 y -1. El valor ideal de esta relación es 1 (ya que G es la relación entre el valor obtenido y el valor deseado de la variable de estado); normalmente, el sistema funciona haciendo grandes estas distancias, con lo que su relación tiende a ser la unidad.

El comportamiento a bajas frecuencias será tanto mejor cuanto más sea la distancia del origen del diagrama (w = 0) al segmento 0, -1. Así, en los sistemas en que el diagrama comienza en K (como ocurre en los de primero y segundo orden, figura 21-23), la relación citada vale K / (K + 1).

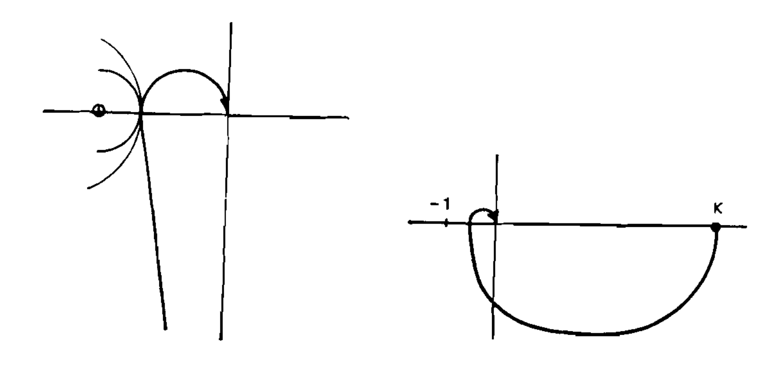


Fig. 21-22. Fig. 21-23.

El error a frecuencia cero será 1/K, que queda como "error residual".

Los sistemas que tienen una integración, como el de la figura 21-17, no tienen error a bajas frecuencias, $G(0) \to 1$, ya que $F \to \infty$ y el error $1/F \to 0$ para $\omega = 0$.

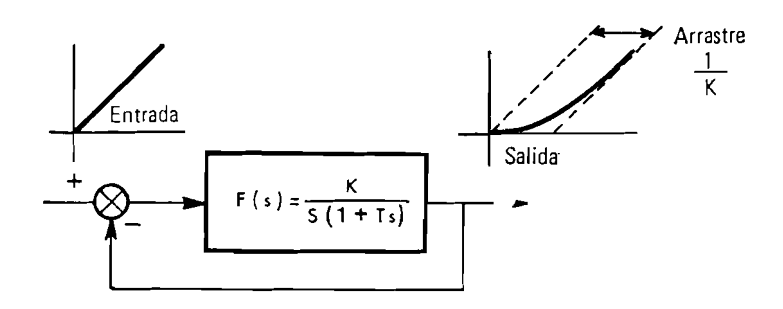


Fig. 21-24.

21.11.- Comportamiento en el tiempo del circuito cerrado.

En el plano de Laplace no es posible hallar el error relativo independiente de la entrada, como se ha hecho para la frecuencia, porque la relación de funciones de transferencia no es la función de transferencia de la relación. El error (absoluto) será, si la entrada es x_0s:

$$E(s) = x_0(s) - x(s) = x_0(s) - x_0(s)\,G(s) = x_0 - \frac{x_0 F}{1+F} = \frac{x_0(s)}{1+F(s)}$$

El valor final del error (**error residual**) en el plano del tiempo se obtendrá por el teorema del valor final, lim s E(s) para $s \to 0$.

Así, si la entrada es el escalón unidad $x_0(s) = 1/s$, el error será:

$$\lim_{s\to 0} s\frac{\frac{1}{s}}{1+F(s)} = \lim_{s\to 0}\frac{1}{1+F(s)}$$

que vale $1/(1 + K)$ si $F(0) = K$ (casos sin integración) y vale 0 si $F(0) \to \infty$ (cuando hay integración). Como se ve, el error residual, que es la desviación final en el plano del tiempo, coincide con el error a bajas frecuencias.

Si se da una entrada en rampa (figura 21-24) $x_0 = 1/s^2$, el error final es:

$$\lim_{s\to 0}\frac{s(\frac{1}{s})^2}{1+F(s)} = \lim\frac{1}{sF(s)}$$

En el caso en que $F(0) = K$, este límite no existe y el error tiende a ∞, pero en el caso de una integración, del tipo K/s en que $sF(s)$ tiene un límite cuando $s = 0$, existe un límite y un error llamado de arrastre que vale $1/K$ (ver figura 21-24).

En general, el comportamiento en el plano del tiempo se puede deducir de las actuaciones en frecuencia con las siguientes reglas prácticas, cualitativas:

a) El **error residual** y la precisión para **"regulación lenta"** son el error residual y la precisión a **bajas frecuencias**.

b) El régimen **oscilatorio** transitorio ocurre a la **"frecuencia propia"** o de resonancia y es más acusado cuanto más aguda es la resonancia; o sea, menos los márgenes de estabilidad.

c) La **rapidez** de regulación viene relacionada (normalmente) con el comportamiento a **altas frecuencias**. Si el sistema es regular, el tiempo de reacción está relacionado con la respuesta de frecuencia, como lo estaba el tiempo de subida con la frecuencia de corte de los amplificadores (inversamente).

21.12.- Compensación de los sistemas de regulación automática.

Se llama compensación al proceso de mejora de una regulación determinada, conseguida fundamentalmente por dos procedimientos:

a) **Compensación integral.**

Consiste en la introducción de un elemento del tipo integral puro $1/s$ en la cadena de regulación, en un sistema que inicialmente no tuviese integración. Como la respuesta de frecuencia es $1/j\omega$, el diagrama se transforma como se indica en

la figura 21-25. Las bajas frecuencias aumentan en módulo, girando $-\pi/2$. Las altas frecuencias disminuyen en módulo, girando también. Como consecuencia se obtiene una respuesta mejor a bajas frecuencias (se gana precisión) perdiendo a altas frecuencias (se pierde rapidez). Dependiendo del sistema particular se podrá o no incrementar ahora la amplificación (según si se ha alejado o no del punto crítico). En general, la frecuencia de resonancia será ahora más baja. También puede incluso ocurrir que aparezca inestabilidad por haberse acercado el diagrama al punto -1, en cuyo caso habría que disminuir la amplificación.

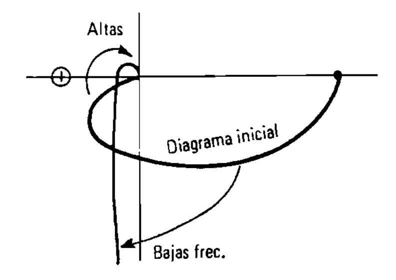


Fig. 21-25.

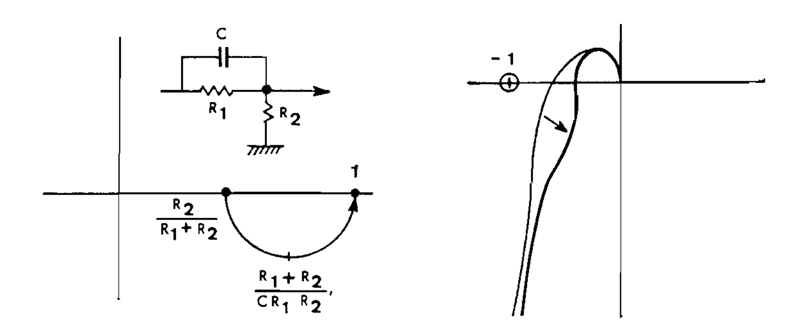


Fig. 21-26. Fig. 21-27.

b) **Compensación diferencial.**

Consiste en alejar el diagrama del punto -1 mediante un **avance de fase**, por medio de un circuito tipo CR como los estudiados en el capítulo anterior. Si se utiliza un circuito como el de la figura 21-26, por ejemplo, se consigue avanzar la fase en las pulsaciones próximas al valor

$$\omega = \frac{R_1 + R_2}{C R_1 R_2}$$

alejándose así del punto -1 como se indica en la figura 21-27. Posteriormente, una vez efectuada esta estabilización por avance

de fase, se puede aumentar la amplificación, ganando así calidad a todas las frecuencias.

Ambos métodos, **integral y diferencial**, se combinan en la práctica a la hora de proyectar el sistema, constituyendo así un primer paso de síntesis del mismo.

SELECCION DE PROBLEMAS RESUELTOS

1.- Dibujar el circuito equivalente (generador con impedancia interna) del circuito de la figura.

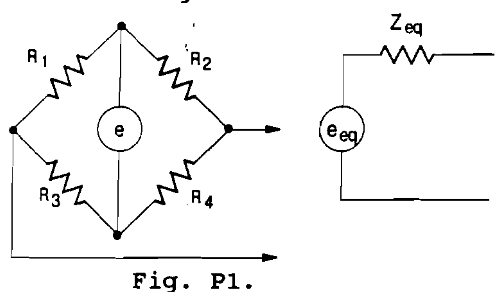


Fig. P1.

Solución.

Tomando como referencia el terminal inferior del generador, aplicamos el teorema de Thevenin a los dos divisores de tensión. Sus tensiones equivalentes e impedancias internas son:

$$e_1 = e\frac{R_3}{R_1+R_3} \quad Z_1 = \frac{R_1R_3}{R_1+R_3} \quad e_2 = e\frac{R_4}{R_3+R_4} \quad Z_2 = \frac{R_2R_4}{R_2+R_4}$$

La tensión equivalente a la salida será la diferencia de tensiones y la impedancia la suma de impedancias, o sea:

$$e_{eq.} = e\left[\frac{R_4}{R_2+R_4} - \frac{R_3}{R_1+R_3}\right] \quad Z_{eq.} = \frac{R_2R_4}{R_2+R_4} + \frac{R_1R_3}{R_1+R_3}$$

2.- Dibujar la forma de onda en el punto A, indicando el valor de máximos y mínimos.

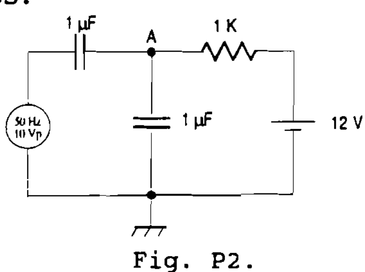


Fig. P2.

Solución.

a) Aplicamos el teorema de Thevenin al divisor formado por los dos condensadores. El circuito equivalente será (mitad de tensión y mitad de impedancia, es decir, doble capacidad):

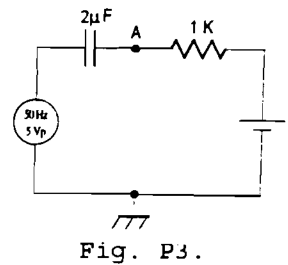


Fig. P3.

b) Aplicamos el principio de superposición. El circuito equivalente para c.a. se obtendrá cortocircuitando la batería (fig. P3):

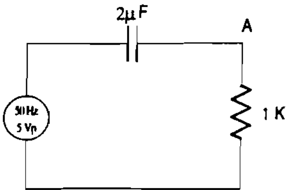


Fig. P4.

Es un divisor CR. La tensión a la salida A será:

$$V_{c.a.}=5\frac{1000}{1000+\dfrac{1}{j\times 2\pi\times 50\times 2\times 10^{-6}}}=5\frac{0,63j}{1+0,63j}=2,65_{[58°]}$$

El circuito equivalente para c.c. se obtiene con los condensadores abiertos, por lo que $V_{c.c.}$ = 12 V.

c) La suma de las ondas será la dibujada en la fig. P5, con los valores indicados de máximos y mínimos:

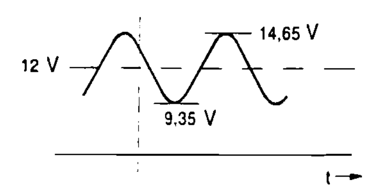


Fig. P5.

3.- Escribir las ecuaciones de la red dibujada en la fig. P6, a)

en el plano del tiempo, b) en el plano de Laplace y c) en el plano jw.

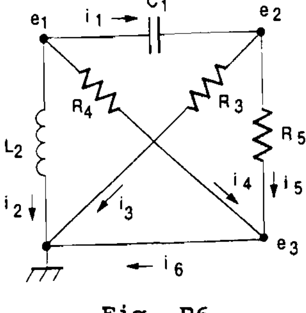


Fig. P6.

Solución.

Las variables de la red son (ver figura) e_1, e_2, e_3, i_1, i_2, i_3, i_4, i_5, i_6.

a) En el plano del tiempo:

$$C\frac{d(e_1-e_2)}{dt}=i_1$$
$$e_1=L_2\frac{di_2}{dt}$$
$$e_2=R_3 i_3$$
$$e_1-e_3=R_4 i_4$$
$$e_2-e_3=R_5 i_5$$
$$e_3=0$$
$$i_1+i_2+i_4=0$$
$$i_1-i_3-i_5=0$$
$$i_4+i_5=i_6$$

b) En el plano de Laplace (las tres últimas son las mismas):

$$e_1-e_2=\frac{i_1}{sC}$$
$$e_1=sL_2 i_2$$
$$e_2=R_3 i_3$$
$$e_1-e_3=R_4 i_4$$
$$e_2-e_3=R_5 i_5$$
$$e_3=0$$

c) En el plano $j\omega$ (las tres últimas son las mismas de a)):

4.- Calcular los parámetros técnicos del cuadripolo de la figura.

$$e_1 - e_2 = \frac{i_1}{j\omega C}$$
$$e_1 = j\omega L_2 i_2$$
$$e_2 = R_3 i_3$$
$$e_1 - e_3 = R_4 i_4$$
$$e_2 - e_3 = R_5 i_5$$
$$e_3 = 0$$

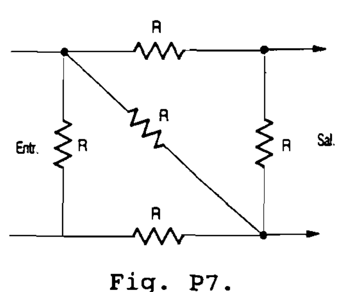


Fig. P7.

Solución.

La impedancia de entrada con salida abierta será:

$$Z_E = \frac{R\,[R + \frac{2R^2}{3R}]}{R + R + \frac{2R^2}{3R}} = \frac{\frac{5}{3}R^2}{\frac{8}{3}R} = \frac{5}{8}R$$

La impedancia de salida con la entrada cortocircuitada será:

$$Z_S = \frac{[\frac{R^2}{2R} + R]\,R}{\frac{R^2}{2R} + R + R} = \frac{\frac{3}{2}R^2}{\frac{5}{2}R} = \frac{3}{5}R$$

La amplificación sin carga será (calculando un primer divisor con carga y después un segundo divisor):

$$A = \Gamma = \frac{R}{\frac{2}{3}R + R} \times \frac{1}{2} = \frac{3}{5} \times \frac{1}{2} = \frac{3}{10}$$

5.- Calcular los valores de la pulsación de resonancia del filtro LC-CL de la figura.

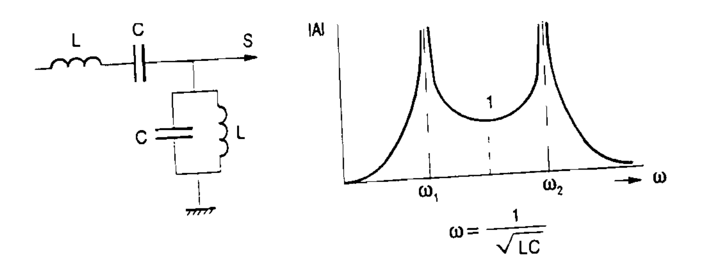


Fig. P8.

Solución.

La respuesta de frecuencia es:

$$A=\frac{\dfrac{\dfrac{L}{C}}{j\omega L+\dfrac{1}{j\omega C}}}{\dfrac{\dfrac{L}{C}}{j\omega L+\dfrac{1}{j\omega C}}+j\omega L+\dfrac{1}{j\omega C}}=\frac{\dfrac{L}{C}}{\dfrac{L}{C}+\dfrac{L}{C}+\dfrac{L}{C}-\omega^2L^2-\dfrac{1}{\omega^2C^2}}$$

$$A=\frac{1}{3-\omega^2LC-\dfrac{1}{\omega^2LC}}\quad \textit{que vale } A=1 \textit{ para } \omega=\frac{1}{\sqrt{LC}}$$

Ahora podemos calcular los polos, que son los valores de ω que anulan el denominador:

$$\omega^2LC+\frac{1}{\omega^2LC}=3 \quad \textit{o sea} \quad (\omega^2LC)^2-3\omega^2LC+1=0$$

$$\omega^2LC=\frac{3}{2}\pm\sqrt{\left(\frac{3}{2}\right)^2-1}=1,5\pm\sqrt{1,25}=1,5\pm1,12,\quad \omega_1^2LC=2,62,\quad \omega_2^2LC=0,38$$

$$\omega_1=\sqrt{\frac{2,62}{LC}}=\frac{1,62}{\sqrt{LC}}=1,62\omega_r,\quad \omega_2=\sqrt{\frac{0,38}{LC}}=\frac{0,616}{\sqrt{LC}}=0,616\omega_r$$

6.- Dibujar la respuesta de frecuencia del filtro de la figura.

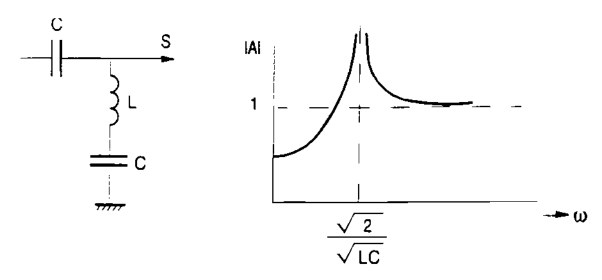


Fig. P9.

Solución.

$$A=\frac{j\omega L+\frac{1}{j\omega C}}{j\omega L+\frac{2}{j\omega C}}=\frac{1-\omega^2 LC}{2-\omega^2 LC}$$

$$Para\ \omega=0,\ A\to\frac{1}{2},\ para\ \omega\to\infty,\ A\to 1$$

$$Resonante\ para\ \omega=\frac{\sqrt{2}}{\sqrt{LC}},\ A\to\infty.$$

7.- Dada la función del tiempo f(t)=cos 30t cos 50t, especificar las componentes del espectro complejo en amplitud.

Solución.

El espectro real del producto contiene las frecuencias (pulsaciones) suma y diferencia, es decir, 20 y 80 rad./s. y amplitud la del producto dividida por dos, es decir 1/2.

El espectro complejo se obtiene del real desdoblándolo sobre el eje de ordenadas, es decir, contiene las pulsaciones -80, -20, +20 y +80, todas con amplitud 1/4.

8.- Resolver la ecuación diferencial y''- 3y'= cos ωt usando la transformación jω.

Solución.

La ecuación transformada da la solución y :

$$-\omega^2 y-3j\omega y=1 \quad y=\frac{1}{-\omega^2-3j\omega}$$

$$|y|=\frac{1}{\sqrt{9\omega^2+\omega^4}}=\frac{1}{\omega\sqrt{\omega^2+9}} \quad \varphi=-\text{arc }tg\ \frac{3\omega}{\omega^2}=-\text{arc }tg\ \frac{3}{\omega}$$

En el plano del tiempo, la solución será:

$$y(t)=\frac{1}{\omega\sqrt{\omega^2+9}}\cos(\omega t-\text{arc }tg\ \frac{3}{\omega})$$

9.- Resolver la ecuación diferencial y'' = f(t) siendo f(t) el escalón unidad, a) en el plano de Laplace y b) en plano del tiempo.

Solución.

a) La ecuación transformada es s^2 Y = 1/s, o sea s^3 Y = 1, Y = 1/s^3. La transformada 1/s^2 corresponde a la función rampa t (integral del escalón unidad) y la transformada 1/s^3 a su integral y = t^2 /2.

b) Directamente:

$$y=\int\int_0^t 1dt^2=\frac{t^2}{2}$$

10.- Hallar la transformada de Laplace de la rampa de la figura.

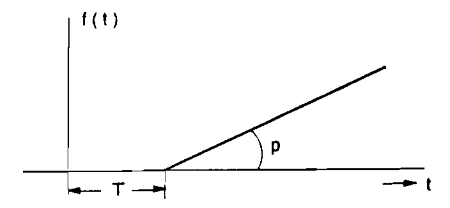


Fig. P10.

Solución.

La rampa de pendiente p se transforma en p/s^2 y aplicando el teorema del retardo se obtiene:

$$Y(s)=\frac{p}{s^2}e^{-sT}$$

11.- El ruido a la salida de un receptor de 40 dB de amplificación es 1,5 Vef. La temperatura de ruido a la entrada es 27ºC, la resistencia de entrada es 100 ohmios y el ancho de banda 1 MHz. Calcular el valor de la señal mínima detectable a la entrada y el número de ruido del receptor.

Solución:

40 dB equivale a A=100 (en amplitud), por lo que R_l = 1,5/100 = 15 mV_{ef} es la señal mínima detectable.

$$R_0=\sqrt{RKT\Delta f}=\sqrt{100\times1,38\times10^{-23}\times300\times\times10^6}=0,64\mu V_{ef.}$$

El número de ruido será:

$$F=\frac{R_1}{R_0}=\frac{15.000}{0,64}=23.437,5$$

$$FdB=20\ \log_{10}23.437=20\times 4,37=87,4dB.$$

12.- Calcular los valores correctos de las resistencias del amplificador de banda ancha de la figura para que la amplificación sea aproximadamente -10 y la impedancia de entrada aproximadamente 10 K. El punto de trabajo del transistor será $I_E \approx I_C$ = 1 mA., V_{BE} = 0,5 V., β = 50. Se supondrá I_{co} = h_{22} = 0. Se tomará V_C = 6 V. en el punto de trabajo.

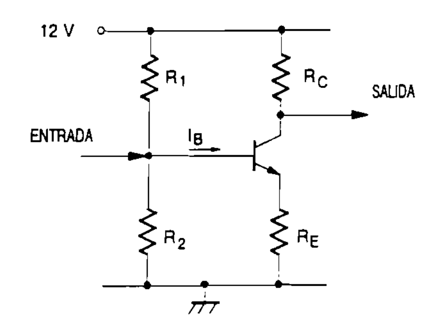


Fig. P11.

Solución.

Al fijar V_C = 6 como criterio de proyecto, se puede calcular R_C = (12-6)V/1mA = 6K. Para que la amplificación sea aproximadamente -10, R_E = R_C / 10 = 600 Ω. La impedancia de entrada del transistor será:

$$Z_E=\beta\,(R_E+\frac{1}{39I_E})=50\,(600+\frac{1000}{39})=31.283\Omega$$

La impedancia de entrada será este valor en paralelo con R_1 y R_2. Como el valor dominante es R_2, fijaremos ese valor en 15K, para que en paralelo con Z_E resulte aproximadamente 10K.

R_1 se fijará para cumplir la condición de punto de trabajo: La tensión (c.c.) en la base será R_E I_E + V_{BE} = 1,1 V. La intensidad por R_2 será 1,1 / 15K = 0,073 mA. La intensidad de base será I_B = I_C /β = 1 mA / 50 = 0,02 mA. La intensidad por R_2 será 0,073 + 0,02 = 0,093 mA, y el valor de R_1 será:

$$R_1=\frac{12-1,1}{0,093}=\frac{10,9}{0,093}=117K.$$

13.- Dibujar la respuesta de frecuencia del amplificador de banda ancha de la figura, especificando niveles y zonas de transición. Los dos transistores funcionan en el punto de trabajo del ejercicio anterior. C_p representa las capacidades parásitas, o un condensador puesto intencionadamente para que el corte final esté bien determinado.

Solución.

La impedancia de entrada de la segunda etapa, antes de que sea apreciable C_p , será (ver ejercicio anterior) el resultado de poner en paralelo los valores Z_E = 31.283, R_1 = 15.000 y R_2 = 117.000, lo que da R_{T2} = 9,32 K.

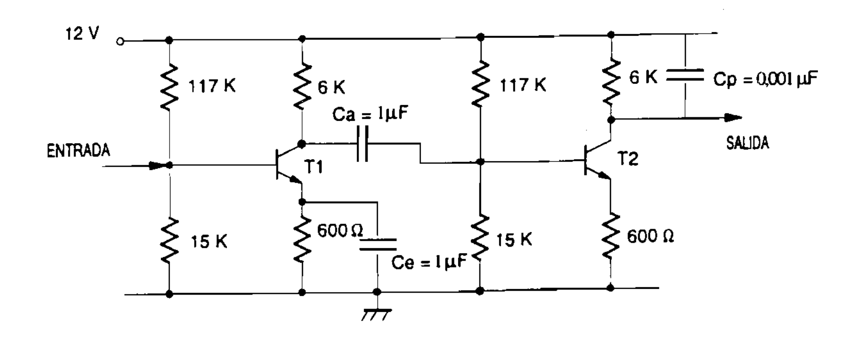


Fig. P12.

Las zonas de transición de los RC del circuito corresponden a las pulsaciones:

$$\omega_A = \frac{1}{9.320\times10^{-6}} = 107,3 \ rad/seg. \ (acoplamiento)$$

$$\omega_E = \frac{1}{600\times10^{-6}} = 1.667 \ rad./seg. \ (emisor)$$

$$\omega_p = \frac{1}{6.000\times10^{-9}} = 166.667 \ rad./seg. \ (C_p)$$

La amplificación será 0 para ω<<107 y ω>>166.667.

El circuito equivalente entre 107 y 1.667 tendrá C_A cortocircuitado, C_E y C_p abiertos. La amplificación de T2 será:

$$A_{T2} = \frac{-0,98\times6.000}{600+\frac{1.000}{39}} = -9,4$$

La impedancia de entrada ya calculada es 9,32K, que está en

paralelo con la impedancia de colector de T1, 6K, con lo que la impedancia de colector que ve T1 será 3,65K. La amplificación de T1 será:

$$A_{T1}=\frac{-0,98\times3.650}{625,6}=-5,72$$

La amplificación total a este nivel será:

$$A_1=(-9,4)\times(-5,72)=53,8\ (34dB).$$

A partir de ω = 1.667 el valor de la amplificación de T2 es el mismo, pero la impedancia de emisor de T1 se anula y la amplificación de T1 y la total pasan a ser:

$$AT1=\frac{-0,98\times3.650}{25,6}=139,7$$
$$A_2=(-9,4)\times(-139,7)=1.313,2\ (62dB).$$

La curva aproximada de respuesta de frecuencia será (no a escala):

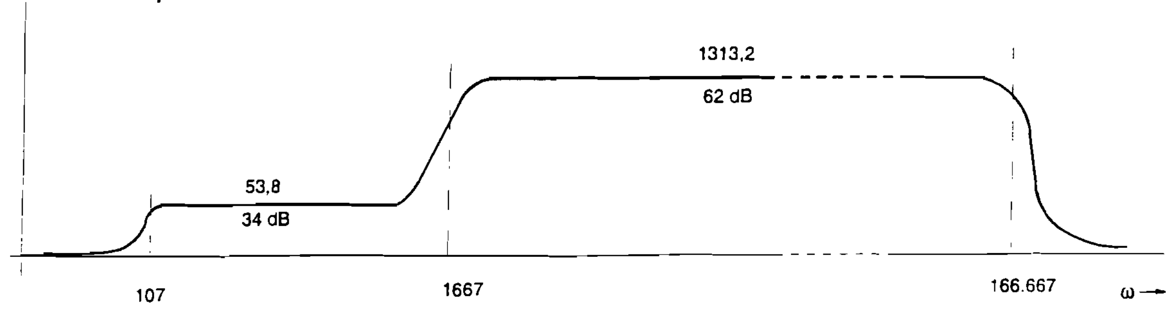


Fig. P13.

La impedancia de entrada es 9,32K (la de T1) y la de salida 6K (la de T2) hasta frecuencias muy altas en que se anula.

14.- El circuito de la figura es un amplificador selectivo, seguido de un seguidor de emisor para bajar la impedancia de salida. Dibujar la respuesta de frecuencia (o de pulsación) calculando la amplificación e impedancias de salida y entrada en resonancia y el ancho de banda. Se suponen los dos transistores correctamente polarizados, T1 en el mismo punto de los problemas anteriores, β = 50 para ambos.

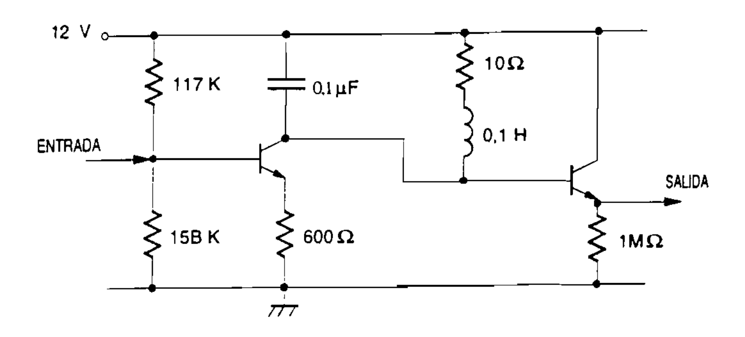


Fig. P14.

Solución.

La pulsación de resonancia es:

$$\omega_r = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{0,1\times10^{-7}}} = 10.000 rad./seg.$$

La Q del circuito resonante, la impedancia en resonancia, la amplificación y el ancho de banda serán:

$$Q = \frac{1}{R}\sqrt{\frac{L}{C}} = 100 \quad Z = Q^2R = 100K.$$

$$A = \frac{-0,98Z}{625,6} = -156,6$$

$$\Delta\omega = \frac{\omega_r}{Q} = \frac{10.000}{100} = 100 rad./seg.$$

La curva de amplificación será:

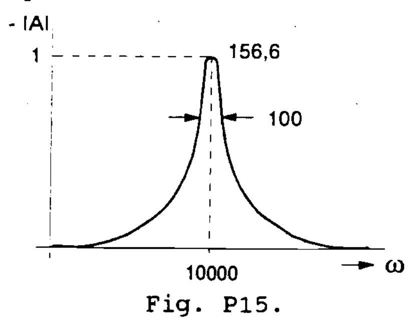


Fig. P15.

El seguidor de emisor tiene amplificación 1 (la curva no cambia). La impedancia de salida se divide por β y resulta ser

100.000 / 50 = 2.000 ohmios. en resonancia (valor máximo).

15.- Calcular la amplificación que debe tener el amplificador de la figura para que el circuito (oscilador con puente de Wienn) oscile, y a qué frecuencia lo hará.

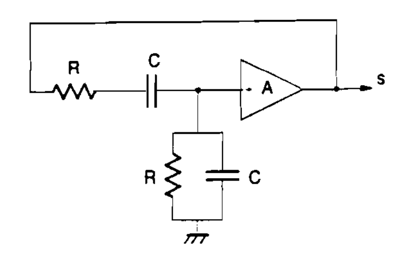


Fig. P16.

Solución.

La amplificación del divisor de tensión es:

$$\beta=\frac{\frac{R}{1+j\omega CR}}{R+\frac{1}{j\omega C}+\frac{R}{1+j\omega CR}}=\frac{1}{3+j(\omega CR-\frac{1}{\omega CR})}$$

El máximo es real y vale 1/3 para pulsación ω = 1 / CR. El circuito oscilará a esa pulsación para A ligeramente mayor de 3 y que disminuya con la amplitud de las oscilaciones hasta cumplir con el valor 3 la condición βA = 1.

16.- Resolver la ecuación diferencial y'' + y' + 10y = x, con valores iniciales y_0, y_0', con un circuito analógico con amplificadores operacionales ideales.

Solución.

Despejamos y'' = x - y' -10y, que cablearemos como aparece en la figura P17.

Los conmutadores cargan los condensadores a los valores iniciales y deben abrirse en el momento 0, a partir del cual se resuelve la ecuación para x(t) aplicada también a partir del mismo momento 0.

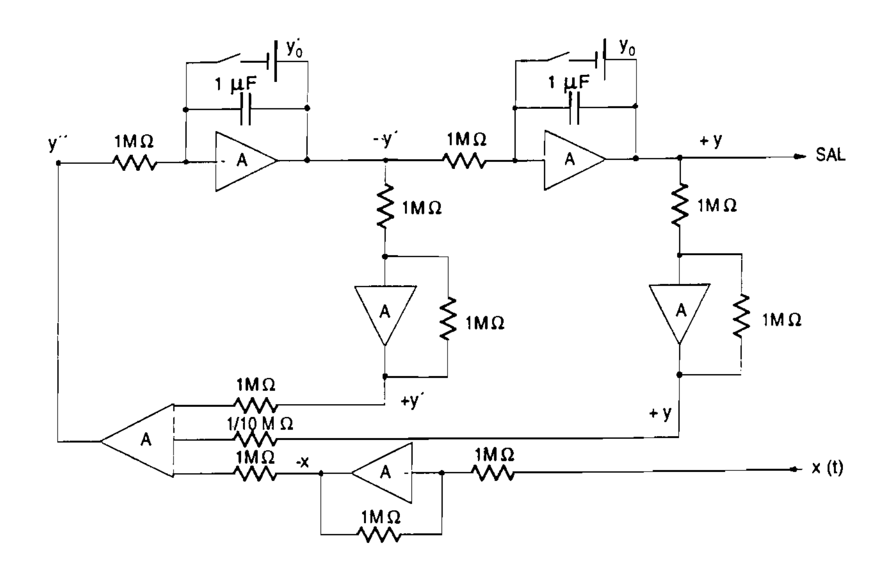


Fig. P17.

17.- Se construye un Schmitt- Trigger con un amplificador operacional como se indica en la figura. Calcular los puntos de conmutación y dibujar el circuito de histéresis. Las tensiones de saturación son de 10 voltios, positivos o negativos.

Solución.

Los puntos de conmutación son:

$$V_1 = 5 + \frac{10-5}{10} = 5,5 \quad y \quad V_2 = 5 + \frac{-10-5}{10} = 3,5$$

El ciclo de histéresis será el dibujado en la figura.

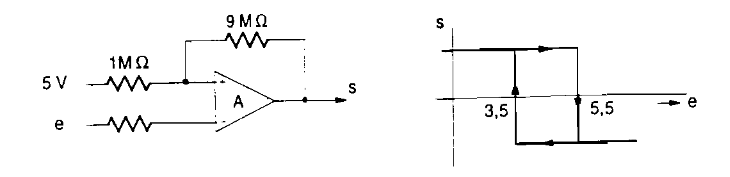


Fig. P18.

18.- Se usa un rectificador de onda completa con transformador de 220:12+12 $V_{ef.}$. La resistencia interna del circuito rectificador es r = 20 ohmios. Calcular el valor "nominal" de la carga para la que la tensión continua es 12V. Calcular el condensador necesario para que la ondulación sea menor del 3% a esa carga.

Solución.

El valor de φ para que E_E sen $\varphi = E_E / \sqrt{2}$ (E_E es el valor de punta) será φ = 45º = $\pi/4$ rad. La ecuación del rectificador nos da ahora el valor nominal de R:

$$ctg\frac{\pi}{4}=\frac{\pi}{2}+\frac{10\pi}{R}-\frac{\pi}{4}=1 \quad R=\frac{10\pi}{1-\frac{\pi}{4}}=\frac{31,4}{0,2146}=146\Omega$$

que corresponden a una intensidad de carga de 12/146=82mA.

La ondulación será (onda completa, τ = T/8):

$$\frac{\Delta E_C}{E_C}=\frac{4\tau}{CR}=\frac{T}{2CR}=\frac{20\times10^{-3}}{2C\times146}=0,03$$

$$C=\frac{10^{-2}}{146\times0,03}=2.283\mu F$$

Se tomará un valor por exceso, por ejemplo, 3.000 μF.

19.- Una fuente de alimentación de c.c. varía entre 11 y 13 voltios (12 V. nominales). Estabilizar con diodo Zener una salida de 8 V. para una carga que varía entre 100 y 200 mA. Se dispone de un Zener de 8 V. con r_Z = 0,5Ω, potencia máxima 10 W., intensidad mínima 25 mA. Calcular entre qué valores debe variar la resistencia R asociada al Zener.

Solución.

A partir de la potencia máxima, la intensidad máxima por el diodo Zener será:

$$I_{Zmáx.} = 10/8 = 1,25 \text{ A.}$$

Aplicando la inecuación que da los límites de R, éstos serán:

$$\frac{E_{Emáx.}-E_Z}{I_{Zmáx.}+I_{Smín.}}<R<\frac{E_{Emín.}-E_Z}{I_{Zmín.}+I_{Smáx.}}$$

$$\frac{13-8}{1,25+0,1}=3,7<R<\frac{11-8}{0,025+0,2}=13,3$$

20.- Tomando una resistencia de 10 ohmios en el ejemplo del problema anterior, dibujar la curva de tensión de salida en función de la intensidad de salida, para E = 12 V.

Solución.

La tensión estabilizada tiene una pendiente de

$$-\frac{Rr_Z}{R+r_Z}=-\frac{10\times0,5}{10,5}=-0,48\Omega$$

que pasa a ser -10 sin estabilización. El Zener deja de estabilizar para un consumo

$$\frac{E_E-E_Z}{R}=\frac{12-8}{10}=0,4A.$$

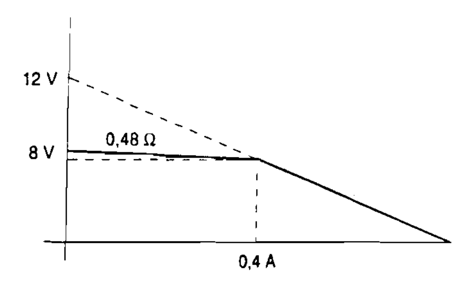


Fig. P19.

21.- Adaptar una línea de Z_0 = 50 ohmios terminada en una impedancia Z_L = 20 ohmios, mediante un fragmento de línea abierta en paralelo, con el diagrama de Smith. λ = 1m.
25
Solución.

La admitancia referida a la impedancia característica es 50/20 = 2,5. Se busca el punto de admitancia 2,5 + 0j. Se describe un círculo a derechas, hacia el generador, hasta encontrar el círculo de admitancia real 1, obteniendo así el arco recorrido de x_1 = 0,09 λ = 9 cm. Se lee la parte imaginaria, -0,95j. Por el círculo y= + 0,95j (el cambio de signo se hace para anular en paralelo la parte imaginaria) se va a la periferia (línea abierta) y se recorre a izquierdas, hacia la carga, hasta llegar a admitancia 0 (línea abierta). El arco recorrido es x_2 = 0,12 λ = 12 cm.

22.- Se controla un vehículo de función de transferencia

$$\frac{1}{(1+T_1 s)(1+T_2 s)}$$

con un controlador integral de función de transferencia K / s. Se supone la medida ideal, $\beta = 1$. Calcular el valor máximo de K para que el control sea estable.

Solución.

La función de transferencia y la respuesta de frecuencia en circuito abierto son:

$$F(s)=\frac{K}{s(1+T_1 s)(1+T_2 s)}=\frac{K}{s+(T_1+T_2)s^2+T_1 T_2 s^3}$$
$$F(\omega)=\frac{K}{-(T_1+T_2)\omega^2+j(\omega-T_1 T_2\omega^3)}$$

La función de transferencia no tiene polos "inestables" (de parte real positiva) pero sí un polo en el origen que habrá que eludir.

El diagrama de Nyquist ($F(\omega)$) corta al eje real en un punto en que la pulsación es aproximadamente la de resonancia. La pulsación de resonancia y el corte se obtendrán así:

$$\omega-T_1 T_2\omega^3=0, \quad \omega^2=\frac{1}{T_1 T_2} \quad \omega_r=\pm\sqrt{\frac{1}{T_1 T_2}} \quad F(\omega_r)=-\frac{KT_1 T_2}{T_1+T_2}$$

Para evitar el polo al aplicar el criterio de Nyquist, se hace $s = \rho\, e^{j\theta}$ (θ varía entre $-\pi/2$ y $+\pi/2$, gira a izquierdas). F es entonces equivalente a $K\, e^{-j\theta} / \rho$, luego gira a derechas, como se ve en la figura.

El criterio de Nyquist dice que el sistema es estable si el diagrama de Nyquist, recorrido con ω de $-\infty$ a $+\infty$ y vuelta por el círculo artificial fijado no da ninguna vuelta alrededor del punto -1.

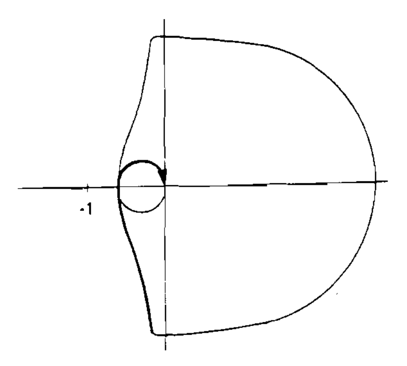


Fig. P20.

La condición para que esto no ocurra es:

$$\frac{KT_1T_2}{T_1+T_2}<1$$

o sea que el valor máximo de K será:

$$K_{máx.}=\frac{T_1+T_2}{T_1T_2}$$

Normalmente se fija K en un valor del orden de 0,7 $K_{máx.}$



BELLISCO
EDICIONES TÉCNICAS
Y CIENTÍFICAS


ISBN 84-85198-77-8